

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (DD-MM-YYYY) 15-10-2004		2. REPORT TYPE Final report		3. DATES COVERED (From - To) Mar 2004 – Aug 2004
4. TITLE AND SUBTITLE Novel Airborne Video Sensors; Super-Resolution Multi-Camera Panoramic Imaging System for UAVs		5a. CONTRACT NUMBER W31P4Q-04-C-R105		
		5b. GRANT NUMBER		
		5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Negahdaripour, Shahriar		5d. PROJECT NUMBER		
		5e. TASK NUMBER		
		5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Braum Image Tech Inc. 600 Puerta Ave. Coral Gables, FL 33143		8. PERFORMING ORGANIZATION REPORT NUMBER BTI-0002		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Adv. Research Proj. US Army Aviation & Missile Command 3701 N. Fairfax Dr. AMSRD-AMR-WS-DP-SB Arlington, VA 22203-1714 Bldg. 7804, Rm 212 Redstone Arsenal, AL 35898		10. SPONSOR/MONITOR'S ACRONYM(S)		
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION / AVAILABILITY STATEMENT The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the Defense Advanced Research Project Agency or the U.S. Government.				
13. SUPPLEMENTARY NOTES High-Resolution Aerial Imagery, Terrain Mapping, Video Servo, Performance Assessment				
14. ABSTRACT Application of a camera array as a flexible, reconfigurable, inexpensive high-resolution panoramic motion-imagery sensor for low-altitude reconnaissance aircrafts is investigated. Assuming multiple-view noisy image position measurements of terrain features and known camera projection matrices by calibration, terrain feature localization and UAV positioning are analyzed by computer simulations, with/without supplementary gyro and GPS. How various system parameters impact the achievable precision of panoramic system in 3-D terrain feature localization and UAV motion estimation is determined for the A=0.5-2 [km] flight altitude range. Enhancement of estimation accuracy from GPS and gyro is explored. Estimation error variance plots are given as a function of camera resolutions, viewing angles, flight altitudes, GPS and altitude measurement errors, number of views, etc. Selected results, from point correspondences in 4[Kpix]x4[Kpix] images and utilizing GPS readings with one-meter error variance at 0.5-2 [km] altitudes, comprise: Estimating 3-D coordinates of ground features tracked in 1-2 dozen images with A/10 baselines at sub-meter accuracy; Determining UAV pose with 0.1-0.3 [deg] variance by matching 2-3 dozens of features in two views. The results provide valuable guidelines for the integration of camera-array images into one super resolution panorama, registering multiple panoramas to construct a single composite view, integration of visual servo with onboard sensors, map-based navigation and AUV positioning. Computed performance charts enable the design of optimal high-resolution imaging system based on the UAV size and capability constraints.				
15. SUBJECT TERMS High-Resolution Aerial Imagery, Terrain Mapping, Video Servo, Performance Assessment				
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT			c. THIS PAGE
		UU	106	19b. TELEPHONE NUMBER (include area code) 786-210-8293

20041108 127

BEST AVAILABLE COPY

"Novel Airborne Video Sensors"

October 15, 2004

Sponsored by

Defense Advanced Research Project Agency (DOD)
Controlling DARPA Office

ARPA Order P372-43

Issued by U. S. Army Aviation and Missile Command Under

Contract No. W31P4Q-04-C-R105

Name of Contractor: *Braim Image Tech, Inc.*

Principal Investigator, Project Scientist, or Engineer: *Shahriar Negahdaripour*

Business Address: *600 Puerta Ave. Coral Gables, FL 33143*

Phone Number: *(786)210-8293*

Effective Date of Contract: *March 1 2004*

Short Title of Work: *Super-Resolution Multi-Camera Panoramic Imaging System for UAVs*

Contract Expiration Date: *October 22, 2004*

Reporting Period: *March 1- August 31, 2004*

DISCLAIMER

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the Defense Advanced Research Project Agency or the U.S. Government.

"Approved for public release; distribution unlimited"

***"Novel Airborne Video Sensors;
Super-Resolution Multi-Camera Panoramic Imaging System for UAVs"***

Abstract

Objective and purpose: Application of a camera array as a flexible, reconfigurable, inexpensive high-resolution panoramic motion-imagery sensor for low-altitude reconnaissance aircrafts is investigated.

Methods employed, restrictions and limits: Assuming multiple-view noisy image position measurements of terrain features and known camera projection matrices by calibration, terrain feature localization and UAV positioning are analyzed by computer simulations, with/without supplementary gyro and GPS.

Results and conclusions: How various system parameters impact the achievable precision of panoramic system in 3-D terrain feature localization and UAV motion estimation is determined for the $A=0.5-2$ [km] flight altitude range. Enhancement of estimation accuracy from GPS and gyro is explored. Estimation error variance plots are given as a function of camera resolutions, viewing angles, flight altitudes, GPS and altitude measurement errors, number of views, etc. Selected results, from point correspondences in $4[Kpix] \times 4[Kpix]$ images and utilizing GPS readings with one-meter error variance at $0.5-2$ [km] altitudes, comprise: Estimating 3-D coordinates of ground features tracked in 1-2 dozen images with $A/10$ baselines at sub-meter accuracy; Determining UAV pose with $0.1-0.3$ [deg] variance by matching 2-3 dozens of features in two views. The results provide valuable guidelines for the integration of camera-array images into one super resolution panorama, registering multiple panoramas to construct a single composite view, integration of visual servo with onboard sensors, map-based navigation and AUV positioning. Computed performance charts enable the design of optimal high-resolution imaging system based on the UAV size and capability constraints.

Contents

1	Summary	1
1.1	Problem	1
1.2	Results	2
1.3	Conclusions	3
2	Introduction	5
3	Methods and Assumptions	6
3.1	Panoramic Views	6
3.2	3-D Reconstruction from Multiple UAV Positions	7
4	Results	10
4.1	Panoramic Imaging System	10
4.1.1	3-D Localization Accuracy:	11
4.1.2	3-D Motion Estimation from 3-D Measurements:	13
4.2	3-D Reconstruction from Multiple 2-D Views	15
4.3	Closed-Form Eight-Point Algorithm	16
4.4	Small Rotation Angle Approximation	17
4.5	Estimation Ground Targets from Nonlinear Method	19
4.6	Pose Estimation; Known Ground Targets	20
	Table 1	21
	Table 2	22
5	Discussions, Conclusions and Recommendations	22
	References	26
	Appendix 1: Theoretical Estimation of Uncertainty Bounds	29
1.1	3-D Reconstruction by Triangulation	29
1.1.1	Accuracy in 3-D Reconstruction	30
1.2	3-D Motion Estimation by Absolute Orientation	31
1.2.1	Accuracy in Motion Estimation	33
1.3	Positioning	35
1.3.1	Accuracy in 3-D Positioning	35
	Appendix 2: 3-D Reconstruction from Multiple 2-D Views	36
2.1	Closed-Form Eight-Point Solution	36
2.2	Non-Linear Solution	37
2.3	Closed-Form Solution with Small-Angle Rotations	38
2.4	Construction of Terrain 3-D Points	39
2.5	Estimation of Motion from Known Target Points	39
	Figures 1-3	40
	Appendix 3a: Simulations of panoramic high-resolution imaging system	43
	Figures 4-12	44
	Appendix 3b: Results of closed-form eight-point algorithm for terrain feature localization	53

<i>Figures 13-15</i>	54
Appendix 3c: Results of closed-form solution with small-angle rotations for terrain feature localization	72
<i>Figures 16-18</i>	73
Appendix 3d: Sample results from nonlinear iterative solution for terrain feature localization	91
<i>Figures 19-21</i>	92
Appendix 3e: Complete results from nonlinear iterative solution for terrain feature localization	98
<i>Figures 22-24</i>	99
Appendix 3f: Results of nonlinear iterative solution for UAV pose estimation	105
<i>Figure 25</i>	106

Lists of figures and tables

Figures:

Figure 1: Schematic diagram of a down-look camera cluster for the construction of conical panoramic views, depicted with 6 cameras as an example. Maps below display the terrain region sizes viewed by various number of cameras in a 12-camera panoramic imaging system, at 3 different UAV elevations of 100, 500 and 2000 meters.

Figure 2: Images showing relative terrain coverage for a fixed camera resolution and field of view at altitudes 100 [m], 200 [m], and 500 [m], and comparison in image resolutions over a terrain feature.

Figure 3: An experiment at 500 [m], showing XY projections of 30 ground features (red circles) imaged from 5, 7, ... , 19 UAV positions (black x). 25% features with highest error variance (blue square) and those 3 times above the median (red dot) are identified, typically located near the boundaries either far away or below a cluster of UAV viewing positions.

Figure 4: Uncertainty in the estimated x component of the terrain feature position at 3 UAV altitudes of 100 [m], 500 [m] and 2 [km] for various system parameters, including number of cameras, camera resolution, and imaging system radius that controls the camera baselines.

Figure 5: Uncertainty in the estimated Z component of the terrain feature position at 3 UAV altitudes of 100 [m], 500 [m] and 2 [km] for various system parameters, including number of cameras, camera resolution, and imaging system radius that controls the camera baselines.

Figure 6: Variation in localization uncertainty when varying the camera slant angle θ based on the number (top) and resolution (bottom) of the cameras.

Figure 7: Uncertainties in the estimation of 3-D motion parameters, rotation $[w_x, w_y, w_z]$ and translation $[t_x, t_y, t_z]$ while varying camera resolution and number of cameras.

Figure 8: Uncertainties in estimating the 3-D translation components, assuming knowledge of rotational motion with an uncertainty of 0.1 [deg], while varying the number and resolution of cameras.

Figure 9: Uncertainties of 3-D motion parameters, rotation $[w_x, w_y, w_z]$ and translation $[t_x, t_y, t_z]$ varying radius of imaging system structure with 12 cameras for various camera resolutions.

Figure 10: Uncertainties of 3-D motion parameters, rotation $[w_x, w_y, w_z]$ and translation $[t_x, t_y, t_z]$ varying density of terrain feature points used in the computations for various camera resolutions.

Figure 11: Uncertainties of 3-D motion parameters, rotation $[w_x, w_y, w_z]$ and translation $[t_x, t_y, t_z]$ with varying the viewing angle of 12 cameras at various resolutions.

Figure 12: Assuming knowledge of rotation with an uncertainty of 0.1 [deg], 3-D translation uncertainties are determined while varying the viewing angle of 12 cameras at various resolutions.

Figure 13: Closed-Form 8-Point Algorithm-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 500); See text for details.

Figure 13: (continued)- Tracking 45 (left) and 60 (right) points with noise-free GPS.

Figure 13: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

Figure 13: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

Figure 13: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

Figure 13: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

Figure 14: Closed-Form 8-Point Algorithm-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 2000).

Figure 14: (continued)- Tracking 45 (left) and 60 (right) points with noise-free GPS.

Figure 14: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

Figure 14: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

Figure 14: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

Figure 14: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

Figure 15: Closed-Form 8-Point Algorithm-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 4000).

Figure 15: (continued)- Tracking 45 (left) and 60 (right) points with noise-free GPS.

Figure 15: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

Figure 15: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

Figure 15: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

Figure 15: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

Figure 16: Closed-Form Solution with Small-Rotation Approximation-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 and 30 (right) points with noise-free GPS (Altitude 500); See text for details.

Figure 16: (continued)- Tracking 45 (left) and 60 (right) points with noise-free GPS.

Figure 16: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

Figure 16: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

Figure 16: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

Figure 16: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

Figure 17: Closed-Form Solution with Small-Rotation Approximation-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 2000).

Figure 17: (continued)- Tracking 45 (left) and 60 (right) points with noise-free GPS.

Figure 17: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

Figure 17: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

Figure 17: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

Figure 17: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

Figure 18: Closed-Form Solution with Small-Rotation Approximation-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 4000).

Figure 18: (continued)- Tracking 45 (left) and 60 (right) points with noise-free GPS.

Figure 18: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

Figure 18: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

Figure 18: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

Figure 18: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

Figure 19: Nonlinear Iterative Solution-- Uncertainty (reconstruction variance [m]) of 3 sample points in terrain feature localization by tracking 3 (left) and 6 (right) points in images with LxL ($L=\{1,3,4\}$) resolutions (Altitude 500); See text for details.

Figure 19: (continued) - Tracking 9 (left) and 12 (right) points.

Figure 20: Nonlinear Iterative Solution-- Uncertainty (reconstruction variance [m]) of 3 sample points in terrain feature localization by tracking 3 (left) and 6 (right) points in images with LxL ($L=\{1,3,4\}$) resolutions (Altitude 2000); See text for details.

Figure 20: (continued) - Tracking 9 (left) and 12 (right) points.

Figure 21: Nonlinear Iterative Solution-- Uncertainty (reconstruction variance [m]) of 3 sample points in terrain feature localization by tracking 3 (left) and 6 (right) points in images with LxL ($L=\{1,3,4\}$) resolutions (Altitude 4000); See text for details.

Figure 21: (continued) - Tracking 9 (left) and 12 (right) points.

Figure 22: Nonlinear Iterative Solution-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 3 (left) and 6 (right) points in images with LxL ($L=\{1,3,4\}$) resolutions (Altitude 500); See text for details.

Figure 22: (continued) - Tracking 9 (left) and 12 (right) points.

Figure 23: Nonlinear Iterative Solution-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 3 (left) and 6 (right) points in images with LxL ($L=\{1,3,4\}$) resolutions (Altitude 2000); See text for details.

Figure 23: (continued) - Tracking 9 (left) and 12 (right) points.

Figure 24: Nonlinear Iterative Solution-- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 3 (left) and 6 (right) points in images with $L \times L$ ($L=\{1,3,4\}$) resolutions (Altitude 4000); See text for details.

Figure 24: (continued) - Tracking 9 (left) and 12 (right) points.

Figure 25: Variances of UAV pose angles- computed from rotation with respect to ref. coordinate system using Rodriguez formula- for various altitudes, GPS measurement uncertainties, and number of terrain feature points tracked in two views. See text for details.

Tables:

Table 1: Variances of feature positions and variations with resolution and GPS uncertainty for 3 altitudes. Some discrepancies, where accuracy is slightly better for a lower GPS accuracy, to the random selection of points in various simulations. See text for more details.

Table 2: Pose angles estimation variances by tracking known terrain targets at various altitudes. See text for more details.

1 Summary

1.1 Problem

Development of automatic image analysis and interpretation systems has been among major activities supported by DoD (DARPA). Video surveillance in high-security areas, potentially followed by automated face identification and recognition, abnormal event detection and mapping are only a few applications of interest that directly involve the extraction of accurate information from images, while covering as large of a scene area as possible. Image resolution is directly tied, first to *the ability* to extract the sought after scene information, next to *the accuracy* of this information, and finally to the *actions that would be triggered* based on the acquired knowledge. One scenario is to locate a person/object in an image, to extract facial/structural features from one or more images accurately enough to recognize the person/target, and to invoke proper action(s) upon recognition (identification or classification) and (or) confidence level of the outcome. Such capabilities require motion video sensors that would generate very high-resolution images of the entire surrounding environment.

Increased resolution can be achieved by constructing a view, ideally a panorama, from the images of a camera cluster each covering a smaller section of the entire field of view [9, 10, 11, 27, 28, 29, 32, 33, 38]. Here, the processing power of the computer is exploited to our advantage to carry out the necessary calculations for the alignment of the images. Existing powerful PCs can achieve a seamless alignment of several standard CCD-resolution images—roughly half to one dozen—in real time. Furthermore, the views can be generated into super resolution imagery [5]; much higher than for each view, though this requires more extensive computations. Such capabilities will improve with the continuous growth in processing power. Furthermore, images from nearby positions (with sufficiently large baselines) enable the application of photo-mosaicing technology to construct rather large composite views [31].

A *conical panorama* can be generated from vertical scans as an oblique down-look camera rotates about a vertical axis, either through or at some distance from the projection center [9]. A slanted down-look configuration is particularly suitable for flyover imagery, while the adjustment of the oblique viewing angle permits the control of the coverage area and overlap in various cameras. One goal of this Phase-1 SBIR project is to analyze the multi-camera realization of a conical panoramic imaging system, targeted for airborne employment at a range of altitudes. Due to flexibility in the choice of certain parameters, the system can be designed to optimize various performance criteria, including the coverage and resolution based on the overlapping regions of neighboring cameras, for a suitable altitude range. Another aspect of the work is to explore the role of image resolution in video servo for UAV positioning over a terrain, say during a surveillance mission. The

objective is to establish charts of performance measures—precisions in 3-D terrain target localization and UAV motion estimation and positioning—in various cases and operational ranges.

To meet the objectives, a number of cases have been studied. In the first half of the project, the emphasis was placed on the assessment of the conical panoramic imaging system, and the performance in terrain reconstruction and motion estimation based on the images from a camera cluster. In the second half, the study emphasized more generally the role of image resolution on performance, independent of the configuration of the cameras in constructing the image; a single high-resolution camera with a large field of view, or a camera cluster each with a lower resolution, smaller field of view, and in arbitrary configurations. Furthermore, attention was given to the fact that the terrain often appears relatively flat in flyover imagery, particularly at high altitudes. As a result, there is generally an inherent translation-rotation visual motion ambiguity that results in some level of inaccuracy in estimating the motion components, namely discerning certain translation and rotation components. To resolve this issue, the use of onboard odometry information, GPS measurements, gyros, etc., can be instrumental. In particular, it is important to know how the use of gyro and GPS measurements may improve visual servo. We have explored the accuracy in determining the 3-D coordinates of terrain features from multiple views along the path, utilizing GPS measurements to establish the UAV's absolute positions. Consequently, we can establish the precision in estimating the UAV pose by tracking/matching the 2-D projections of located landmark features at different UAV positions. While gyros can provide an estimate of the UAV orientation, the drift error can become significant in long-duration surveillance operations. Thus, integration of visual servo with gyro can potentially provide a mechanism to estimate the drift, as well as enhance the total performance.

1.2 Results

In assessing the performance of the multi-camera panoramic imaging system, we have utilized well-known solution for 3-D scene reconstruction [17] and motion estimation [3, 20, 21], but have derived mathematical models to assess the accuracy of these solutions in terms of various system design parameters. We have carried out computer simulations to establish variations in system performance with these parameters, and provided charts of performance -vs- system parameters. This is helpful to determine the pay-off in the number and resolution of individual cameras.

The application of three different methods has provided the assessment of accuracy in the estimation of 3-D motion and terrain target positions from multiple views. Numerous computer simulations with these techniques have been carried out while varying image

resolution, UAV altitude, GPS uncertainty, number of views, and image feature localization uncertainty. Charts of error variance have been generated to readily analyze achievable accuracy under various cases, and to establish the pay-off in increased resolution at various altitudes.

1.3 Conclusions

For the conical imaging system, increased image resolution can enhance 1) 3-D localization accuracy of the terrain targets, 2) estimation of the UAV's motion by utilizing the correspondences of terrain features at nearby viewing; 3) computation of the UAV position along UAV's trajectory for registering images taken at multiple views to construct a large composite image. Due to the small baseline (separation) of cameras within the imaging system structure, ability of depth perception from the visual cues in multiple cameras is limited to very low flying altitudes, say at terrain distances of less than $Z = 100$ [m]. In this range, depth perception can be substantially enhanced, to within 2-3 meters uncertainty, by the deployment of many (order of 1-2 dozens) very high-resolution (e.g., QSXGA and QUXGA at 6-8 [MPix]) cameras. In contrast, high level of X and Y accuracy—say localization uncertainty of roughly a decimeter—is feasible by the deployment of half a dozen average resolution VGA cameras. Use of more high-resolution cameras can reduce the uncertainty down to a couple centimeters. Slanted viewing angle of the cameras can provide not only larger coverage, but surprisingly some improvement in localization accuracy of terrain features and targets. Motion estimation accuracy is affected somewhat adversely by the well-known inherent translation-rotation ambiguity of visual motion, and the uncertainty in the estimation of depth values from the camera cluster. Integration with high-precision angle sensors—e.g., ring laser gyros (RLG)—can substantially improve the performance, to the point where the translation motion uncertainty is only limited by the gyro inaccuracy.

The altitude limitation and the somewhat inaccurate depth reconstruction accuracy is readily overcome by the use of multiple images from relatively large baselines. Each image may be generated with the multi-camera panoramic imaging system, or alternatively may be acquired with a single high-resolution camera (distinction is immaterial in assessing the accuracy variation with resolution). We only need to assume the image resolution within the desired field of view. Furthermore, use of GPS measurements lends several advantages. First, GPS readings with baselines of 100 [m] or more between any two UAV views provides a more accurate estimate of the UAV absolute position than visual motion techniques, and furthermore is drift free. Incorporating this information enables us to establish the UAV pose along its trajectory from visual cues. Finally, integration with gyro measurements offers a mechanism to estimate and rectify gyro drift in long-duration operations, and to

compute pose more accurately by data fusion.

To provide some numbers, with multiple images at 1) $4[\text{Kpix}] \times 4[\text{Kpix}]$ resolution; 2) extending over roughly 90° in field of view; 3) GPS error variance of $\sigma_{GPS} = 1$ [m], and altitude measurement variance of $10\sigma_{GPS}$ [m], we can achieve sub-meter terrain feature localization accuracy in each coordinate: An error variance as low as $0.1\text{--}0.9$ [m] by tracking a few dozen features over a couple (or more) dozen views with a baseline $1/5\text{--}1/10$ of the altitude at $500\text{--}4000$ [m]. We can also determine the UAV orientation with error variances of less than $0.15\text{--}0.25^\circ$ in pitch and roll and $0.01\text{--}0.05^\circ$ in heading at $500\text{--}4000$ [m] altitudes, by tracking two dozens of features with GPS error variance of $\sigma_{GPS} = 1$ [m] and UAV altitude uncertainty of $15\sigma_{GPS}$ [m].

GPS and altitude uncertainties both play a critical role in the accuracy of the UAV pitch and roll motions, but have much less impact on determining the heading. Both measurement errors are less significant at higher altitudes (as a percentage), thus leading to better localization and pose estimation accuracies. The performance can be improved by increasing the field of view, reducing the UAV altitude measurement error, in addition to algorithmic adjustments and optimizations that have not been considered in this study.

Before moving forward, we make a terminology clarification. When talking about camera or robot “pose” in the vision literature, one often refers to both the position and orientation. The work in the second half of this study primarily deals with the estimation of the UAV orientation from visual cues, when GPS estimates are used to establish the UAV absolute position. Thus, “pose estimation,” “estimated pose,” “variance of pose parameters,” etc., primarily refers to the UAV orientation. Unless it clearly means otherwise from the context, UAV pose means the orientation at some position along its path.

2 Introduction

An image is a two-dimensional (2-D) view of a three-dimensional (3-D) world from a specific viewpoint. An image acquired by a traditional camera has limited resolution and field of view. An evolving imaging technology is based on the emergence of a new paradigm, the so-called “*computational camera*”. This refers to a digital device that embodies the unification of imaging sensors and the computer. Based on projection geometry and image remapping, various capabilities become feasible that are either impossible, too complex, and (or) very costly to achieve with traditional cameras and lenses. Examples include the computer processing of typical images (projections onto planes) to construct views that correspond to 1) any arbitrary camera view, or 2) projections onto arbitrary image surfaces, say spheres, cylinders, and cones. Significant technological advantages are being made possible by exploiting computing power that is growing at a tremendous rate and diminishing cost. This enables carrying out more and more complex operations that result in higher and higher image quality. However, one still needs devices that can provide views of the entire, or some relevant large region of the, scene. As depicted in fig. 1, a cluster of N cameras properly positions relative to each other— each covering some region of the scene— can provide a panoramic coverage. The alignment of the images to generate the panorama is performed based on well-known theories in camera calibration and image registration [18]. The computational speed of current powerful PCs can enable the registration of several (order of half to one dozen) images at video rate, enabling the construction of panoramic imagery as quickly as the camera images are recorded.

The number and resolution of each individual camera control the resolution of the constructed panorama. Clearly, the highest quality image is expected when utilizing a larger and larger number of cameras with the highest possible resolution. However, this has to be traded off against the requirements for data storage, processing, transmission, etc. The goal of this study is to assess the achievable accuracy and resolution in terms of the total number, resolution, and configuration of individual cameras. The accuracy and resolution are tightly related and their relevance is tied to the application. In the first half of our work, we explored the impact of a number of design parameters of the panoramic imaging system on terrain feature localization.

In this second half of our work, we have been primarily concerned with the accuracy in locating certain ground landmark targets, and consequently tracking them to determine the aircraft’s pose from recorded images and GPS readings. Unlike the first half of the project, where the investigation also addressed the impact of camera cluster configuration within the panoramic system, the emphasis here is on the image resolution, UAV altitude and GPS measurement uncertainty. In particular, we have explored two operational scenarios:

1. The UAV is circling over some terrain, say a metropolitan area, in a long surveillance

operation. It has no prior knowledge of the terrain. By locating and tracking certain fixed targets over several images and utilizing GPS readings, it computes its position and pose, and more importantly the 3-D positions of the tracked targets.

2. Based on knowledge of some fixed ground landmarks—e.g., determined by the method in the first scenario, the aircraft establishes its position and (or) trajectory. While GPS gives the position and gyros provide pose information, gyro measurements do drift over an extended time. Thus we seek the utilization of a drift-insensitive vision-based approach, which may simultaneously allow the estimation and (or) correction of gyro drift, as well as an effective estimation method potentially based on the integration of gyro and visual servo solutions.

The information determined above can feed into the system for building a super resolution map/mosaic of the terrain, and/or to track certain moving targets of interest.

3 Methods and Assumptions

3.1 Panoramic Views

Reconstruction of a terrain map from multiple images acquired on an airborne system has been known to photogrammetry engineers since the early 1900's [4, 8, 30]. Over the last two to three decades, the same mathematical models have been deployed extensively in the computer vision community for numerous applications, including the automated operation of robotics systems, mobile platforms and mapping systems. In analyzing the panoramic imaging system, we have utilized well-known solutions for 3-D scene reconstruction [17] and motion estimation [3, 20, 21], however, we have incorporated new mathematical models that we have derived to assess the sensitivity to measurement noises for various system parameters. These models enable us to assess the accuracy in localizing a terrain feature based on multiple observations in the camera cluster as a function of various system parameters—error in determining the image position of a terrain feature, number and resolution of the cameras, viewing angle, altitude, and area coverage. (Appendix 1 gives the detailed theoretical foundation of the results given here.) In particular, determination of feature positions as image measurements is tied to the quantization level in a digitized image; the higher the resolution, the more accurate the feature position. Charts of accuracy-vs- imaging system parameters have been generated to readily analyze and assess the performance of a multi-camera panoramic imaging system, and to establish the pay-off in the number and resolution of individual cameras, their arrangement and viewing angles, etc.

In our simulations, the primary assumption is that each camera satisfies the pinhole camera model [17]. In practice, this requires the calibration of each camera 1) to determine its internal parameters, including the focal length, aspect ratio (horizontal to vertical scaling ratio) and the actual location of the image center (where optical axis pierces the image plane); 2) rectifying each image to remove lens distortions [18]. Furthermore, we assume perfect knowledge of the relative positions and orientations of the cameras in the coordinate frame of the panoramic imaging system. In practice, this is determined by a variety of external calibration methods [35, 37]. Typical achievable accuracy, measured in terms of the average misalignment error of features used in calibration, is some small fraction of a pixel size, which is sufficient for most applications.

3.2 3-D Reconstruction from Multiple UAV Positions

Application of three different methods—detailed in Appendix 2—has provided the assessment of accuracy in the estimation of 3-D motion and terrain target positions from multiple views. The underlying constraint is the rigid body motion constraint

$$P_j = R_{ij}P_i + t_{ij}$$

where P_i denotes the coordinates of a terrain point in coordinate system of the camera at position i , and $\{t_{ij}, R_{ij}\}$ are 3-D vector and 3×3 rotation matrix that describe the displacement and angular motion of the aircraft from position i to j . It is well-known that estimation of both translational and rotation motions from image projections of features on a relatively flat scene¹ can be an ill-conditioned problem [1]. The use of GPS provides several advantages. First GPS and accurate altitude measurements with baselines of 100 [m] or more between two UAV positions generally provide better estimates of the UAV absolute position than visual motion techniques, while being drift free. This simplifies the problem to estimating the UAV pose from visual cues. Furthermore, integration of this information with gyro measurements offers a mechanism to estimate and rectify gyro drift in long-duration operations, and to compute pose more accurately through data fusion. Therefore, we determine displacement vector t_{ij} between two UAV positions T_i and T_j , from the combination of GPS and altitude measurements up to some varying levels of accuracy:

$$t_{ij} = T_j - T_i.$$

We will discuss the impact of GPS and altitude measurement errors as factors influencing the system performance.

¹In our case, the terrain appears flat when viewed from the aircraft at high altitude, since height variations due to building and other structures are typically negligible relative to the aircraft altitude.

A large number of simulations have been carried out for each of the two scenarios described in the introduction section, while varying a number of parameters:

- Image resolutions of $L [Kpix] \times L [Kpix]$ ($L = \{1, 3, 4\}$);
- Flight altitude A [m] ($A = \{500, 2000, 4000\}$);
- Number of views M ($M = \{2 : 21\}$);
- Number of terrain features N (typically $N = \{15, 30, 45, 60\}$, though other cases have also been tested for certain algorithms);
- GPS X and Y position variances of $\sigma_{GPS} = \{0, 1, 2\}$ [m];
- Altitude measurement error with variances of $10\sigma_{GPS}$ (in terrain feature localization) and $15\sigma_{GPS}$ in pose estimation;
- Pixel localization error with variance σ_f ($\sigma_f = \{1/3, 1\}$).

We have assumed a lower altitude measurement error variance of $10\sigma_{GPS}$ [m] in feature localization experiments than $15\sigma_{GPS}$ [m] for pose estimation. The motivation here is that generally, we want to establish the locations of terrain features with high accuracy as they are used in subsequent operations for UAV positioning. Therefore, we may use sensors with higher precision in the first scenario, and a more typical sensor in the later operation.

We have assumed a visual field of view (FOV) of roughly 90 [deg] in both horizontal and vertical directions, in all experiments. This gives us a relatively large FOV which is desired for robust motion estimation. The various resolutions within this fixed FOV may be achieved with one or more cameras². As one scenario, the three resolutions we will compare can be realized with a single one-Mpix camera, and 3×3 and 4×4 arrays of cameras, each with one Mpix resolution but a smaller FOV. The FOV was purposely fixed in all the cases, so we can establish merely the impact of image resolution. In particular, since many studies point to improved motion estimation accuracy with larger FOVs, we did not want this fact to corrupt our conclusions.

The three assumed altitudes are treated as low (500 [m]) and intermediate (2000-4000 [m]) ranges. As we will note from the results, error variances for higher image resolutions converge at intermediate altitudes, enabling us to infer the behavior at higher altitudes. For GPS readings, the error-free case study is useful to establish the estimation

²One consideration may be the trade-off in the cost of a high-quality lens with small distortion for a wide-angle FOV (as used in photogrammetry applications) in comparison to lower quality inexpensive lenses and complexity of 1) lens distortion correction with a wide-angle FOV, versus 2) intrinsic/external calibration of 2-3 cameras with smaller FOVs and lower distortions.

error due solely to image feature position inaccuracies³. The range of $M = \{2 : 21\}$ views provide a fairly conclusive assessment of the impact on 3-D target reconstruction accuracy in the first scenario, which would subsequently be utilized for UAV localization in the second scenario.

The feature position uncertainty of $\sigma_f = 1/3$ [pix] is justified by the fact that many feature detection and localization methods perform at sub-pixel accuracy [14, 7, 5]. Specifically, our assumption means that the N selected features have been localized at no worse than $3\sigma_f = 1$ [pix] uncertainty. One can conservatively identify, match and track two orders of magnitude or more features than a few dozens (15-60, used in our study) with the assumed accuracy level, given the abundance of features in 1-16 Mpix images of textured ground scenes. The point selection can be readily accomplished by the application of some robust statistics including RANSAC-based algorithm [13]. Therefore, one interpretation of our assumption is that the matched features have previously passed a threshold test of no worse than one-pixel localization error by a RANSAC-based method⁴. In one case though, for the lower altitude of 500 [m], the results have been derived for $\sigma_f = 1$ [pix], as a more severe test of performance.

The coordinate frame of the imaging system at one view has been chosen as the reference system. In the simulations, the UAV positions were chosen over the terrain area that is seen by the cameras at the reference view. This corresponds to an area roughly $2A \times 2A$ in the XY plane (A is the altitude). The baseline is of the order of $(1/5 - 1/10) * A$ in simulations involving a smaller number of views. Some baselines between pairs of UAV positions become smaller as the number of views increase, say to a dozen or more. Fig. 4 shows selected sample examples from our simulations, where the (red) circles show the terrain feature points, and the (blue) crosses are the UAV positions. We have assumed a frame-to-frame rotational motion in the order of 1-2 [deg] when testing the small-angle approximation method, and a few degrees for all other cases.

In the first scenario described in the introduction section, we have tested three algorithms described in Appendix 2; classical eight-point algorithm [24], nonlinear optimization method, and closed-form solution for small rotation angles. In the second scenario—determining UAV pose by tracking terrain targets—the nonlinear method of section 2.5 in Appendix 2 has been applied. Given the very large number of simulations carried out, it is difficult to establish the best way to arrange the graphs involving the variations in numerous parameters. As an example, *the results from the eight-point algorithm, for the*

³In the remainder, unless implied or stated otherwise, GPS readings refer to measurements of all three components of UAV position, thus including the altitude measurements. However, reader should note that we have assumed a different level of uncertainty for the two measurements; see above.

⁴This is a reasonable assumption, given the size of our images. Also, we have computed the feature noises from a Gaussian distribution randomly, and thus some pixels may have a higher noise than $3\sigma_f = 1$ [pix].

altitude of 500 [m] while varying other parameters, have been tabulated in 6 pages each with 120 plots. We have presented the figures for various simulations in Appendices 3(a-f), and have given a summary and general conclusions in the main body of the report.

As indicated above, the terrain points have been selected randomly. Each simulation was repeated 300 times with different noise samples– for errors in image positions, GPS readings, and imprecise knowledge of terrain feature positions in the second scenario– in order to determine the estimation statistics, namely the variance as the uncertainty measure. In some cases, the camera positions and viewing angles are not ideal for certain points, resulting in poor estimates of their positions. In practice, such situations can be readily identified by examining the reprojection errors. However, the scope of this work has not been to develop a suitable implementation of some 3-D reconstruction algorithm, but rather the investigation of performance. As a result, we have selected to eliminate as outliers 25% (out of $N = \{30, 45, 60\}$) of the points with the highest reconstruction variances. (This means that the variance plots show the performance for $N' = 3N/4$ points.

For a sample experiment, fig. 4 depicts the projection onto the XY plane of the randomly selected terrain features (red circle) and the camera positions (black x). We have not shown the Z coordinates of these features, (heights above ground reference plane), which have been generated with a mean and variance of 8 [m]. The UAV altitudes have been varied, from nominal $\{500, 2000, 4000\}$ [m] values, by about 10-20 [m] in experiments with smaller number of views, and as large as 50-75 [m] when approaching the maximum number of views (say, 15-21 views). In these plots, certain terrain features are marked with red dots, and others with a blue rectangle. These, respectively, correspond to 1) points with reconstruction accuracy above 3 times the median $3 * med$, and 2) $N - N'$ points that have been discarded from our results. This example shows that with smaller number of views, the discarded points have variances larger than $3 * med$, while some discarded points have a variance less than $3 * med$ as the number of views increase. Also, these points are typically clustered near the region boundaries, where there is little variations in viewing direction from several AUV positions.

4 Results

4.1 Panoramic Imaging System

The images in fig. 1 depict the terrain size covered by various number of– from a total of N – cameras in the panoramic imaging system, at altitudes of $A = 100$ [m](top-right), $A = 500$ [m](bottom-left) and $A = 2000$ [m] (bottom-right). In these images, $N = 12$ cameras have been assumed, each with a resolution of 1024×768 , roughly 70 [deg] in FOV,

and configured at a viewing angle of $\theta = 75$ [deg] (measured from down-look position). The significance of these particular plots is that, by determining areas of the terrain that are imaged by a relatively large number of cameras, we can establish the terrain region size that may be imaged at the highest resolution is a super-resolution panorama (to be performed in the second-half of the project).

In most applications, we are primarily concerned with the central region of the image which maps the area directly below the UAV. We want to view a larger portion of this region with as many of the N cameras as possible. In these plots, we see that when the UAV flies at 100 [m], a central area of 30 [m] \times 30 [m] is imaged by all 12 cameras. This area increases to roughly 150 \times 150 [sqm] for an altitude of 500 [m], and over 500 \times 500 [sqm] for an altitude of 2000 [m]. Unfortunately, with increased elevation, each image pixel corresponds to a larger local region of the terrain, leading to lower resolution. Depicted in fig. 3 are sample 1024 \times 768 images from altitudes 100 [m], 200 [m] and 500 [m]. A measure of relative resolution is the image size of a particular feature in these images. For any UAV altitude A , N such images from the cameras that view different regions of the terrain can be fused to generate a super-resolution panorama. Roughly speaking, 4 cameras at altitude A can readily generate a super-resolution image equivalent to a single image at half the altitude. With respect to our example, a super resolution comparable to the image at 100 [m] can be generated from 4 images at 200 [m], or from a 5 \times 5 camera array at 500 [m]. Clearly, performance can be improved substantially if we start with very high-resolution cameras, say each at 6-8 mega pixels.

4.1.1 3-D Localization Accuracy:

A number of sample experimental results, given in Appendix 3a, demonstrate the application of our analytical results is assessing the performance of a multi-camera imaging system for target localization and mapping. Referring to fig. 4 in Appendix 3a, the top row shows the variance (uncertainty measure) in determining the X position of a terrain feature, detected in various views of the panoramic imaging system. The assumed uncertainty arises from the errors in the image positions of a feature due to image quantization. The X and Y position uncertainties are the same, and thus only X direction results are given. In these plots, we are comparing the accuracy of a system with $N = 6 : 16$ cameras, when flying at three different altitudes of 100 [m], 500 [m] and 2000 [m]. Each curve corresponds to a certain number of cameras. The cameras are arranged symmetrically on a circular ring of radius $d = 0.25$ [m], at a slant angle of 75 [deg]; refer to fig. 1 for system structure. We have shown in the left column the uncertainty measures when the image resolution varies from 640 \times 480 (for a typical VGA camera) to 3200 \times 2400 (for a very high-resolution QUXGA camera). As an example, consider the top-left figure for the 100 [m] altitude. We can determine the X coordinate of a terrain feature with a variance of

less than 10 [cm] when viewed by six VGA cameras. The uncertainty drops to about 6 [cm] when 16 cameras are used, and to about 2 [cm] with 6 high-resolution QUXGA cameras. The corresponding results for a higher 500 [m] altitude show larger uncertainties, roughly 0.47 [m], 0.3 [m] and 0.1 [m] respectively. For an altitude of 2000 [m], we can achieve an accuracy of roughly 1.2 [m] with 16 typical VGA cameras, while 6 high-resolution QUXGA cameras would give a one-sigma position uncertainty of 0.4 [m].

The plots on the right column show little/no advantage in increasing the camera separations within a larger imaging system structure. From these plots, we conclude: 1) We can generally establish the X and Y positions of the terrain feature with relatively high accuracy; 2) We can also determine how the precision varies with various parameters.

As depicted in fig. 5, the story is quite different for position along the Z direction, which is tied directly to the ability to determine the vertical size (or height above ground) of terrain features. At a flying altitude of 100 [m], for example, the estimation uncertainty is as large as 38 [m] when utilizing 6 VGA cameras. This drops to about 27 [m] and 23 [m] with 12 and 16 VGA cameras, respectively. Taking advantage high-resolution digital cameras (e.g., 4-8 [Mpix]), the performance increases significantly, approaching an uncertainty of 10 [m] or better. For 8 [Mpix] QUXGA cameras, the uncertainty is within 6 [m]. The variances are meaningless for the altitudes of 500 [m] and 2000 [m] (most values are larger than the flight altitude). This indicates no ability for 3-D depth estimation with the multiple images of the camera cluster. This very inferior performance (compared to X and Y uncertainty) is not surprising, and is directly tied to the small/negligible baseline of the imaging system (distance between nearby cameras). The performance plots in the right column, with 12 cameras, show lower uncertainties by increasing the imaging system structure radius d to provide a larger separation between the cameras; refer to fig. 1. For example, the uncertainty drops from 27 [m] to roughly 15 [m] or less, if the radius d of the 12-camera imaging system is increased from 0.25 [m] to 0.5 [m] or larger. Another important conclusion of these results is that the percentage improvement is less when high-resolution cameras are used (e.g., roughly from 6 [m] to 4 [m] with 12 QUXGA cameras).

How large the structure's radius can be made is tied to the UAV size and payload capabilities. For example, smaller UAVs would require a more compact low-weight camera cluster assembly, say one with a radius in the order of 0.2 [m] or less. Mid-size UAVs can carry larger units, say 0.2 [m] to 0.35 [m] radii, while the size and weight constraints are less limiting on a larger UAV. Therefore, one direct application of this graph is that for a given UAV platform, we can readily determine how many cameras and what resolutions are necessary to achieve a desired level of accuracy in 3-D depth perception. These results suggest that acceptable accuracy in depth perception is generally restricted to low altitudes. For 3-D estimation and reconstruction from high altitudes, we need to utilize methods

based on multiple view geometry, well-known in the photogrammetry and structure from motion/stereo literatures [4, 8, 17, 30].

An interesting result is the slight variation in accuracy by changing the camera viewing angle θ ; refer to fig. 1. Shown in fig. 6(top) is the localization uncertainty, while varying the number of cameras at 1024×768 resolution. On the bottom, we have fixed to number of cameras to 12, but have changed the resolution. For example, the top plot shows that the 27 [m] uncertainty with 6 cameras at a slant angle of 75 [deg] can be lowered to about 19 [m], if the viewing angle is adjusted to 60 [deg]. We also conclude from these plots that the choice of the viewing angle is more critical when using a smaller number and (or) lower resolution cameras.

4.1.2 3-D Motion Estimation from 3-D Measurements:

Given the estimated 3-D positions of terrain targets from two view points, we can deploy the solution to the absolute orientation problem from photogrammetry to determine the motion of an imaging system from one viewpoint to the next. The problem has a closed-form solution, as given in section 1.2 of Appendix 1 [3, 20, 21]. A number of problems can be addressed based on this solution, including the construction of 2-D photo-mosaics and 3-D terrain reconstruction from panoramas at multiple views. We are interested to establish how accurately the 3-D information can be determined from the images of the multi-camera panoramic system, and how the precision compares with alternative methods. Based on the earlier results— particularly the high inaccuracy in determining the Z component of a terrain feature position from higher altitudes— the following results are determined for only the low altitude of 100 [m]. We have utilized a 10% density of image points (i.e., one out of every 10 pixel) for 3-D motion computations in all of the results given in Appendix 3b, but one case where we analyze the impact of the number of terrain features. Furthermore, these uncertainties result from the errors in image feature positions, and thus hold for any motion size. Therefore, the larger the motion size, the larger the SNR of the motion estimation process.

Fig. 7 depicts the uncertainty in the six translational and rotational components of the 3-D motion as the image resolution of each camera is varied. Each plot gives the variation of these results with the number of cameras. One immediate conclusion is that camera resolution, more than the number of cameras, is a critical factor. At highest image resolution, we roughly have 1 [deg] and 2 [m] uncertainties in the X and Y components of rotation and translation, respectively. In contrast, we have nearly 0.25 [deg] and 0.6 [m] for the Z components, respectively. At first, this appears surprising as our earlier results showed the X and Y components of terrain feature positions to be much more accurate than the Z component. Therefore, given the same level of accuracy in two views, we expect the X and Y (translation) motion components, in comparison to the Z component,

to reflect the same higher accuracy. Why this is not the case is that embedded in these uncertainty results is also the well-known X/Y translation-rotation ambiguity of motion vision. Roughly speaking, a X/Y translation of magnitude T with respect to a target at distance A along the viewing direction and a Y/X rotation of $\tan^{-1}(T/A)$ [rad] produce very similar image motions (or disparities). Thus, erroneous X/Y translation motions (of magnitude T) can be offset by corresponding erroneous Y/X rotations of $\tan^{-1}(T/A)$ [rad]. Referring back to fig. 7, if the X/Y translation has a variance of roughly 2 [m] (when utilizing 12 QUXGA cameras) with respect to a target at about 100 [m], it can be offset with a Y/X rotation with variance $\tan^{-1} 0.02$ [rad] ≈ 1 [deg]. This can be readily verified from the XY rotation uncertainties for 12 cameras (top-left and middle-left plots for 12 QUXGA cameras). As another test, we have a variance of 12 [m] in X/Y translation with 6 low-resolution VGA cameras. This can be offset with Y/X rotation variance of $\tan^{-1} 0.12$ [rad] ≈ 7 [deg], which is the calculated uncertainty for these rotations.

Overcoming the rotation-translation ambiguity of visual motion typically calls for an imaging system with a large field of view, and terrain features with disparate depth values. While the multi-camera conical imaging systems accommodates a relatively large field view, the inability to accurately estimate the depth Z of terrain features works against us. One solution, quite appropriate for UAVs, is to determine the rotational motions from high-precision gyros (e.g., RLG) that are typically installed on these platforms. With rotation known, the translational motion can then be estimated with more precision. Determining translation from the closed-form solution of absolute orientation, with known rotation, is trivial: 1) Apply the known rotation to the measurements in one coordinate system, 2) determine the centroid of the two measurements, and 3) determine the difference of the two centroids. Having established that we can determine the X and Y coordinates of terrain features with high accuracy (order of decimeter or less depending on the number of cameras and resolution), we expect the precision to carry over to the estimated X and Y components of the translational motion. Computer simulation presented in fig. 8 verifies the improvement in the estimated translation, assuming 0.1 [deg] angle measurement variance from the RLG. Included in the new results, with the revised X/Y translation uncertainties of roughly 1 [m], is the component due to rotational uncertainty. (Based on our analysis above, 0.1 [deg] X/Y rotation uncertainty can lead to a translation uncertainty of roughly $100 \tan 0.1 \approx 0.17$ [m] at a 100 [m] terrain elevation.) For consistency with the results in fig. 7 and to restrict our study to the performance of the panoramic imaging system, we concentrate in the balance of this section on the uncertainty measurements where all the motion estimates are determined from visual cues (assuming no information from external angle measuring devices). As we see shortly, the case where the camera angle is varied requires special attention.

In fig. 9, we have determined the variations in the estimation uncertainty when the

radius of a twelve-camera structure is increased. It decreases roughly by a factor of two, when radius changes from 0.25 [m] to 0.6 [m]. We can readily verify that the major contributor to the uncertainty is the translation-rotation ambiguity.

In fig. 10, we have explored how the uncertainty change with the density of the 3-D measurements. On the horizontal axis, "Distance between 3-D points" of 5, 10, ... means that one out of 5, 10, ... points on the target terrain have been utilized in the computations. Increasing the density of the measurements, there is a one-to-one relationship with improvement in the estimation accuracy. In other words, if twice as many points are used, the uncertainty drops by a factor of 2.

Finally, we examine in fig. 11 the impact of the camera viewing angle. Here, there appears to be a contradiction with earlier results, at first glance. While we had established that increasing the camera angles resulted in some slight increased uncertainty in the estimation of the target feature positions, the motion parameters estimated from these 3-D points seem to be more accurate. The explanation goes as follows: When we increase the camera angles, approaching 90 [deg], all cameras look (almost) downward and thus a larger region of the terrain is viewed by nearly all of the cameras. As a result, there is a much lower uncertainty in the estimated 3-D positions of a large portion of the terrain features, compared to the case where the cameras are arranged at a slanted viewing angles. Consequently, we arrive at a more accurate motion estimate when a much larger percentage of the 3-D positions are known with high precision. *This emphasize the advantage in utilizing our theoretical model that accounts for all of the key factors.* We have repeated the same experiment, assuming again knowledge of the rotation angles up to 0.1 [deg] uncertainty; see fig. 12. We note that the X/Y translation uncertainties drop by roughly a factor of 2, down to nearly 0.5 [m] for a viewing angle of 90 [deg].

4.2 3-D Reconstruction from Multiple 2-D Views

We now present the results of various simulations from the later part of our study: To establish accuracies in estimating the 3-D coordinates of terrain features and the UAV poses from images acquired from multiple large-baseline views, utilizing noisy GPS and altitude measurements to determine the UAV's absolute positions. First, 3-D coordinates of selected terrain targets are determined by processing images at different UAV positions along the flight path, say roughly every A/k ($k = 5 : 10$) meters, where A is the flight altitude. Three methods escribed in Appendix 2– closed-form eight-point algorithm, closed-form small-rotation solution, and an iterative method based on non-linear constraints– are tested. The results deal more generally with the role of image resolution on performance, independent of the configuration of the cameras in constructing the image; whether we have a single high-resolution camera with a large FOV, or an array of

cameras as in our panoramic system— each with a lower resolution and smaller field of view— is immaterial. Next, the located features are tracked in image pairs to determine the UAV pose. As stated, gyros can give the same information, but the drift error can be significant in long-duration surveillance operations. Integration of visual servo with a gyro can be considered as a mechanism to estimate and rectify the drift, while enhancing the estimation accuracy.

4.3 Closed-Form Eight-Point Algorithm

The classic eight-point algorithm, described in section 2.1 of Appendix 2, requires (not surprisingly) a minimum of eight point correspondences in two views for a solution. Our simulations have been performed with $N = \{15, 30, 45, 60\}$ points over $M = 2 : 21$ views. In particular, we have been able to establish the improvement with the use of $M > 2$ frames. In doing so, the two-view algorithm was applied to the data (image features) at a reference frame and a second position, estimating the rotation between the two views (using GPS readings to estimate the translation). This was repeated for up to 20 image pairs from a total of 21 (one at reference and 20 other) views. These estimated motions— from the reference view to any one of 20 different views— have been used in applying the solution in section 2.4 of Appendix 2, in order to 1) estimate the 3-D positions of the terrain targets, 2) compare the estimated accuracy of tracked target features for $M = 2 : 21$ views. The results of the simulations for this section are found in Appendix 3b.

As an example, fig. 13 is a sample set of results for the *altitude of 500 [m]*. In each page covering one GPS variance selection, 120 figures depict the uncertainties with varying number of views, image resolutions, and 2 choices for number of points (a second set gives the same for two other choices of number of points). The 20 cases of $M = \{2 : 21\}$ views have been given as a 5×4 array of plots for reconstruction variance -vs- point number. Thus, each page contains 6 such arrays, where 2 side-by-side ones deal with two choices of number of points (e.g., $N = \{15, 30\}$ in one page and $N = \{45, 60\}$ on the next page). Going down the rows, we cover variations in image resolution $L = \{1, 3, 4\}$ [Kpix].

Before moving forward, we need to emphasize that the following conclusions are based on the application of the eight-point algorithm, and not the best accuracy that can be achieved with alternative methods including those discussed in subsequent sections. We start with the noise-free GPS readings, solely to concentrate on impact of image resolution and pixel localization accuracy:

- Estimation uncertainty generally improves with the number of tracked features, views and resolution;
- Variation with the number of views, from 2 to 21, is significant: By a factor of 5-10 with 15 tracked points , and 3-4 with 60 features;

- Improvement with resolution, from 1 to 16 mega-pixel images, is very significant by a factor of 10 or more;
- Reconstruction variance is 1.5-2 [m] at $L = 1$ [Kpix] resolution, going down to 0.1-0.15 [m] at $L = 4$ [Kpix].

The addition of GPS measurement errors skew the above conclusions significantly, both in terms of achievable accuracy and the improvements with the number of tracked features, views and resolution. This leads to the conclusion that, in comparison to the GPS measurement noises, image feature detection inaccuracies have a far less impact on the estimation errors. The simulation results for GPS variances of $\sigma_{GPS} = 1 - 2$ [m] show that

- Estimation uncertainty generally improves with the number of tracked features and views, but not image resolution;
- Variation with the number of views from 2 to 21 is roughly by a factor of 2, when tracking 15-60 points;
- *Reconstruction variance is about 15 [m] and 80 [m], at $\sigma_{GPS} = 1$ [m] and $\sigma_{GPS} = 2$ [m], respectively, based on roughly 40 (out of 60) tracked points over 21 frames at $L = 1 - 4$ [Kpix] image resolution.*

We recall that the key objective in the application of the current solution is to locate selected 3-D terrain targets, and *the simulations were done primarily to assess performance* (this is also why we have not addressed the accuracy of the estimated rotations from the eight-point algorithm). Our results clearly indicate that these accuracies are highly insufficient to provide a suitable mechanism for the second application, which is to track known terrain targets for computing the UAV pose (and/or to estimate the gyro drift). Finally, construction of super-resolution imagery is highly unlikely with such high inaccuracies. Consequently, results for altitudes 2000 [m] and 4000 [m] have been performed and documented in Appendix 3b (figs. 14-15), but we do not elaborate on them here as the next two methods provide far superior performances than the eight-point algorithm.

4.4 Small Rotation Angle Approximation

The method of section 2.3 in Appendix 2 has been employed to estimate the UAV pose and the 3-D positions of ground targets, but it is the latter that we are primarily concerned with in the first operational scenario. We need to clarify that this method typically works well, either 1) where frame-to-frame rotations are small, of the order of a few degrees, or 2) if applied when one frame is de-rotated using some reasonable knowledge, up to a

few degrees, of the underlying (potentially large) rotation angle. This can be based on the (potentially rough) gyro measurement or the image-based estimate from the last UAV position. While minimum 3 points are sufficient to arrive at a solution, conclusions from the experiments with the eight-point algorithm motivated us to start with a higher number of matched features; $N = \{15, 30, 45, 60\}$ as in the previous case.

Fig. 16 in Appendix 3c shows sample results for an *altitude of 500 [m]*. Again, simulations with noise-free GPS readings are informative in assessing the impact of the image resolution and feature detection accuracy:

- Estimation uncertainty improves little with the number of tracked features, more with increasing views and most significantly with image resolution;
- Variation with the number of views, from 2 to 21, is by a factor of about 3-4;
- Improvement with resolution– from 1 to 16 mega-pixel images– is by a factor of 8-10;
- Reconstruction variance is roughly 1.5-2 [m] at $L = 1$ [Kpix] resolution, down to 0.1-0.15 [m] at $L = 4$ [Kpix] resolution.

This GPS noise-free performance is practically the same as with the eight-point algorithm. Not surprising, both methods perform well when the GPS errors vanish. However, GPS measurement errors skew the above conclusions to a much lesser extend than the previous solution. We conclude from the simulation results for the GPS variances of σ_{GPS} of 1-2 [m]:

- Estimation uncertainty generally improves with the number of tracked features and views, as well as with image resolution when tracking features over a small number of views;
- Variation in accuracy, for 2 to 21 views, is roughly by a factor of 2-3;
- Accuracy improves by a factor of 2 from $L = 1$ [Kpix] to $L = 3$ [Kpix] resolution, with no further improvement at $L = 4$ [Kpix] resolution.
- *Reconstruction variance is roughly 3 [m] for $\sigma_{GPS} = 1$ [m], and 15 [m] for $\sigma_{GPS} = 2$ [m] for about 40 (out of 60) tracked points over 21 frames at $L = 1 - 4$ [Kpix] image resolution.*

This performance with noisy GPS measurements corresponds to a six-fold improvement with respect to the 8-point algorithm.

For the intermediate altitudes of 2000-4000 [m] (shown in figs. 17-18 of Appendix 3c), the same general conclusions are drawn. As a result, we simply list the quantitative performance results:

- $A=2000$ [m]: 3-D localization error variance is roughly 0.2 [m] for noise-free GPS data. It jumps to roughly 20 [m] with GPS variance of $\sigma_{GPS} = 1$ [m], and 70 [m] for $\sigma_{GPS} = 2$ [m] at $L = 4$ [Kpix] resolution (for about 40 out of 60 tracked points over 21 frames);
- $A=4000$ [m]: 3-D localization error variance is roughly 0.8 [m] for noise-free GPS data. It increases to roughly 25 [m] with GPS variance of $\sigma_{GPS} = 1$ [m], and finally about 80 [m] for $\sigma_{GPS} = 2$ [m] at $L = 4$ [Kpix] resolution (for about 40 out of 60 tracked points over 21 frames);

4.5 Estimation Ground Targets from Nonlinear Method

We now present the results from the application of the nonlinear method in section 2.2 of Appendix B, to estimate both the UAV pose and the 3-D positions of the ground targets (see Appendix 3d). Again, we are solely concerned with the accuracy of the 3-D targets.

Fig. 19 in Appendix 3d depicts the results of various simulations for the *altitude of* $A = 500$ [m]. For each case, we have shown here *sample results for 3 of the* $N = \{3, 6, 9, 12\}$ *points tracked*. Complete results for all of the points have been given in Appendix 3e. We have used less features that in prior two cases for several reasons: Jumping ahead, the results seem to exhibit asymptotic convergence to a minimum around $N = 12$ points. Furthermore, tracking 12 points provides sub-meter terrain localization accuracy, which we had aimed for. Furthermore, these simulations based on an iterative nonlinear algorithm have been rather time-consuming and become more so as the number of features increase. This imposes(ed) a constraint on how far we would carry the analysis, while also investigating the performance of the two alternative solutions (we started with this solution first, and later tested the other two). As an example, typically a few hundred iterations of the non-linear algorithm need to be carried out for convergence. Furthermore, each simulation has to be performed several hundred times or more, to establish statistical performance measures. In some cases, more has been and is necessary as we have to randomly choose noise samples for the image feature positions and the GPS readings. Finally, these experiments have been repeated for various resolutions and number of frames, in addition to GPS noise levels. While the accuracy may be lowered by increasing the number of terrain features, the asymptotic behavior suggests that the improvement may be relatively marginal.

Various plots have been arranged in 3×3 arrays, rows corresponding to one of the 3 points, while columns are assigned to X , Y and Z components. For each altitude, the results for tracking $N = \{3, 6, 9, 12\}$ points have been arranged on two pages, $N = 3/6$ on one page in the left/right column, followed by $N = 9/12$ points in the next page in the left/right column. On each page, 3×3 arrays are arranged row-wise for different image

resolutions; $L = 1, 3, 4$ (e.g., first-row 3×3 arrays in fig. 19 correspond to 3 of the tracked points with $N = 3$ in the left and $N = 6$ in the right for resolution 1 [Kpix] \times 1 [Kpix], second-row for 3 [Kpix] \times 3 [Kpix], and last row for 4 [Kpix] \times 4 [Kpix]).

Once used to this arrangement convention, we can concentrate on each plot which shows three (R,G,B) curves for GSP uncertainties of $\sigma_{GPS} = \{0, 1, 2\}$. The vertical axis, as before, is the estimation uncertainty, while the horizontal axis represents the number of frames used to track the feature; the more frames, the higher the accuracy. Given the randomness of the process and that the results do vary with the distribution of the image features, there are some variations in this trend, and the dashed line is the smooth approximation of the expected behavior.

From these results, we conclude the following:⁵

- Estimation uncertainty improves when increasing the number of views, tracked features, resolution and altitude;
- Estimation uncertainty is slightly lower in the Z direction;
- The uncertainty is typically less than one meter, when tracking the features over 15 images.
- The GPS inaccuracy is less significant as the number of features and views increase.

The statistics summarizing the complete results, given in figs. 22-24 of Appendix 3e for the 3 altitudes, have been given in table 2. Small discrepancies, where accuracy is marginally better for a lower GPS accuracy, is mainly due to the random selection of points in various simulations: It is conceivable that some cases involved an unfavorable arrangement of feature points—some clustered within a local region of the image, or covered a smaller field of view (as we discussed in presenting the results from the earlier two methods). The impact becomes more evident when less number of features are utilized as in these simulations, in comparison to the previous methods.

However, the main conclusion should be based on the order of magnitude of these numbers, considerably lower than the results based on the closed-form solutions with a larger number of tracked features⁶. Thus, it is possible to achieve sub-meter accuracy with one to two dozen of features, ideally distributed over as large of a FOV as possible, tracked in one to two dozen frames.

⁵Recall that the results for an altitude of 500 [m] have been determined for a feature localization uncertainty of $\sigma_f = 1$, in comparison to $\sigma = 1/3$ for the other 2 altitudes, as a more stringent test.

⁶The higher inaccuracy of the eight-point algorithm is not surprising where the depth variations among the scene points are small. Additionally, we cannot incorporate knowledge from the GPS readings until the essential matrix $E_{ij} = [t_{ij}]_{\times} R_{ij}$ has already been calculated.

Altitude [m]	GPS Err σ_{GPS} [m]	XY Uncertainty [m]			Z Uncertainty [m]		
		1 Kpix	3 Kpix	4 Kpix	1 Kpix	3 Kpix	4 Kpix
500	0	0.7	0.8	0.9	0.3	0.4	0.3
	1	1.1	0.9	0.9	0.3	0.3	0.3
	2	1.7	1.3	0.9	0.3	0.3	0.3
2000	0	0.3	0.3	0.2	0.1	0.1	0.1
	1	0.5	0.3	0.3	0.1	0.1	0.1
	2	1.	0.4	0.4	0.1	0.1	0.1
4000	0	0.4	0.2	0.1	0.1	0.1	0.1
	1	0.4	0.2	0.2	0.2	0.1	0.1
	2	0.5	0.4	0.4	0.1	0.1	0.1

Table 1: Variances of feature positions and variations with resolution and GPS uncertainty for 3 altitudes. Some discrepancies, where accuracy is slightly better for a lower GPS accuracy, is due to the random selection of points in various simulations. See text for more details.

4.6 Pose Estimation; Known Ground Targets

Applying the nonlinear optimization method described in section 2.5 of Appendix 2, we have determined the camera pose by tracking a number of terrain targets whose positions are known up to some uncertainty. We have adopted the achievable accuracy in locating ground targets from the results of the previous section at each altitude, in addition to the assumed GPS uncertainty and image feature localization errors, to perform simulations for determining the error variances of the UAV pose. The angles are expressed as components of a 3-D rotation vector ω computed from the rotation matrix R by the Rodriguez formula; see section 1.2 of Appendix 1. We have investigated the variations with the number of terrain features and the image resolution. Drawing from the results, given in fig. 25 collectively for the three altitudes:

- Estimation error variance decreases with increasing image resolution;
- Improvements in error variance with image resolution grow with GPS error: that is, there is more improvement in accuracy from increasing the image resolution, when GPS measurements are less accurate.
- Estimation uncertainty improves with the number of features tracked. It is unstable when minimum of 3 features are used, as it highly depends on the configuration of these points. It levels off at around 30 features;
- Accuracy in estimating the heading angle is much higher than for pitch and roll (XY) components. In fact, it approaches that for noise-free GPS readings with increased

Altitude [m]	GPS Err σ_{GPS} [m]	ω_{xy} uncertainty [deg]		ω_z uncertainty [deg]			
		1 Kpix	3 Kpix	4 Kpix	1 Kpix	3 Kpix	4 Kpix
500	0	0.12	0.08	0.04	0.07	0.05	0.03
	1	0.39	0.34	0.25	0.18	0.14	0.05
	2	0.71	0.63	0.52	0.32	0.21	0.11
2000	0	0.05	0.02	0.01	0.03	0.01	0.01
	1	0.35	0.31	0.20	0.05	0.03	0.02
	2	0.53	0.40	0.40	0.06	0.06	0.04
4000	0	0.05	0.02	0.01	0.03	0.01	0.01
	1	0.20	0.15	0.15	0.03	0.01	0.01
	2	0.36	0.31	0.30	0.06	0.02	0.02

Table 2: Pose angles estimation variances by tracking known terrain targets at various altitudes. See text for more details.

resolution and number of tracked points. In contrast, the pitch and roll accuracy with erroneous GPS readings never approaches the noise-free case. This is expected due to the impact of *GPS* uncertainty in determining the UAV translational motion, subsequently affecting the interpretation of rotations due to the translation-rotation ambiguity;

- GPS error becomes less significant at higher altitudes.

Table 2 summarizes the variances for various altitudes, image resolutions, and GPS uncertainties. *With GPS error variance at $\sigma_{GPS} = 1$ [m], we can achieve roughly 0.15-0.25 [deg] in XY and 0.01-0.05 [deg] in Z components of the rotation ω for the image resolution of 4 [Kpix] \times 4 [Kpix].*

While we have not reported results here, our limited simulations indicate that the estimation error variances are rather sensitive to the uncertainty in the UAV altitude measurements. Thus, *improvement can be significant with the use of high-precision altimeters.*

5 Discussions, Conclusions and Recommendations

The so-called *computational camera* is undoubtedly a very powerful and flexible paradigm to achieve a seamless unification of imaging sensors and the computer. It accommodates the capability to overcome various complexities and (or) shortcomings of existing imaging systems, with flexible reconfigurable arrangement of several low-cost conventional cameras, in order to generate super resolution imagery, in theory over the entire sphere of viewing directions. Computational powers of today's processors currently forbid real-time

realization, but will be possible in near future with the tremendous growth of processors at diminishing cost.

This work primarily concentrated on the performance assessment of a panoramic system in a range of design parameters— e.g., resolution and number of cameras— that can provide real-time high-resolution imaging and (or) transmission to a central command station, based on technologies available currently or the very immediate future. In fact, performance results based certain parameter ranges in our analysis— say, 1-4 [Kpix] \times 1-4 [Kpix] image resolution over 90 [deg] in FOV— may be viewed as somewhat conservative.

For the earlier part of the work— studying the performance of the panoramic system for terrain imaging— theoretical models of feature localization and motion estimation accuracies have been derived and utilized. We have used as the primary benchmark the ability to localize a terrain feature, and studied how the localization uncertainty varies as a function of a number of imaging system parameters. The results have several implications in the application of the multi-camera technology for aerial imagery. First, they provide quantitative measures on the trade-offs in the number and resolution of the cameras. Next, they enable designing a system that provides optimal performance for a given altitude range. Also, we can conclude what information can be extracted with/without acceptable accuracy. For example, X/Y coordinates (relative to the UAV) can be established with high accuracy from multi-camera system at one view, and variations in X/Y accuracy can be significant with the number of cameras. In contrast, given that the camera separations in a cluster is limited by the UAV size, we cannot extract from the multiple overlapping images useful information about the vertical distance (Z coordinate). Thus, we cannot determine the vertical size of terrain features from the panorama at one view. Yet, a super-resolution panorama can be generated by the integration of multiple images based on a simplistic locally planar terrain model, and we can evaluate the improvement in resolution according to our quantitative results based on the number of cameras. Furthermore, it has been shown that estimate in the vertical direction can be determined reliably by the application of motion vision techniques, given several panoramas (at assumed resolution and field of view) from multiple (dozen or more) UAV positions along its trajectory. Specifically, error variance of 0.3 [m] or less can be achieved, in the 500-4000 [m] altitude range studied.

In assessing the imaging system's performance in computing UAV motions from estimated 3-D coordinates of terrain features at each view, we have utilized the variances of the estimated motion parameters based on a close-form solution to the absolute orientation problem. We have verified significant improvements with increased camera resolution, number of cameras, density of terrain features used for motion estimation, as well as the viewing angle. More importantly, however, utilizing high-precision gyros to determine the UAV pose at each view will result in the most significant impact, allowing us to estimate the translational motion components with much better accuracy. The use of 3-D fea-

ture coordinates from each view to compute 3-D motion with acceptable accuracy has the following drawbacks:

- Limited to very low altitudes, say 100 [m];
- Requires a very large density of features;

Alternatively, we studied the use of GPS readings along the UAV flight path in determining the 3-D coordinates of terrain features. Subsequently, 2-D visual cues by tracking the features across images enables us to estimate the UAV pose. This approach based on 2-D image projections of terrain features works well for higher altitudes while having to track only about 3 dozens of points.

We can list some remedies for improvements based on our findings, but did not investigate them further in this study, in part due to project time limitations and in part for not combating to meet imposed accuracy bounds or limits:

- Positioning performance at lower altitudes can be made better by revising a "simplistic" assumption made in our simulations: We have used the computed error variances of the terrain feature positions at each altitude (section 4.5) to determine the variances of pose parameters for the same altitude. However, given that we can determine the 3-D coordinates of target features more accurately at higher altitudes (because GPS errors are less significant as a percentage of the altitude), 3-D terrain coordinates may be first determined from the images at higher altitudes (e.g., $A = 4000$ [m] or higher) and subsequently used to more accurately estimate pose at lower altitudes.
- Higher resolution at lower altitudes for pose estimation, though the accuracy for the intermediate (and higher) altitudes appears to be leveling off at the resolutions we have explored;
- Establishing UAV positions more accurately, as limited GPS and altimeter accuracies are key factors in higher estimation variances;
- Locating landmarks and fixed terrain targets with better precision using other sources.
- Integration of gyro and visual servo should not only produce better results, but enable the estimation and overcoming of gyro drift.

Certain algorithmic modifications are likely to enhance performance. For example, the later part of our work on target positioning and UAV pose estimation is based on the knowledge of 2-3 dozens of 3-D positions in one coordinate system (reference view) and the 2-D projections in two views. However, 2-D correspondences of numerous other

image features, without knowledge of their 3-D coordinates, can provide more constraints in determining the pose parameters (epipolar constraint of the multiple view geometry, discussed in Appendix 2). For very low altitudes, say 100 [m] or less, the binocular cues from multiple cameras of the panoramic imaging system may be utilized. Finally, terrain relief is often negligible relative to the UAV altitude at mid to high altitudes, and thus the surface may be treated as a flat plane. For this case, closed-form solution for planes can sometimes provide a more accurate estimate [25, 6, 17]. We have not tested the solutions for planar surfaces because we have concentrated on solutions that impose no restriction on terrain topography.

Having said these, a number of factors, not the target of this study, can result in some deterioration in actual performance. For example, we have not directly considered inaccuracies associated with system calibration. As the image resolution increases, determination of the imaging system internal and external parameters become increasingly important. For example, since we can achieve the same resolution over a fixed FOV in several ways, there are certain trade-offs in 1) achieving high-precision external calibration of a cluster of cameras at moderate cost, each with lower FOV and negligible distortion; 2) intrinsic calibration of a single or a smaller number of cameras at higher resolutions and distortion rates; 3) Avoiding performance degradation due to calibration errors by improving the unit camera performance, say distortion rate, at higher cost. Some of these questions may be answered through computer simulation, e.g., assuming higher variances for 2-D image feature positions and (or) a spatially varying uncertainty model that would represent the errors due to image distortion effects.

In conclusion, we have tried to study a number of key issues that play a role in the performance of a high-resolution panoramic imaging system under certain UAV operational conditions. The underlying issue is to establish if image resolution need or should be increased, and the pay-off, to meet certain levels of accuracy. We believe that the results do generally provide a good assessment for the performance of a high-resolution vision system. However, given the number of variable factors, some of which not addressed in this study, the design and simulations of a system should directly take into account the specific requirements of the application (e.g., image resolution), limitations (e.g., weight, size, UAV capabilities), desired accuracies, etc. Consequently, one can address trade-off issues while simultaneously designing conservatively to overcome unaccounted factors.

References

- [1] G. Adiv, "Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field," *IEEE PAMI*, Vol 11(5), May, 1986.
- [2] X. Armangue, J. Salvi, "Overall view regarding fundamental matrix estimation," *Image and Vision Computing*, 21, 2003.
- [3] K.S. Arun, T.S. Huang, and S.D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE PAMI*, Vol 9(5), September 1987.
- [4] K.B. Atkinson, *Close Range Photogrammetry and Machine Vision*, Whittles Publishing, Scotland, 1996.
- [5] D.P. Capel, "Image mosaicing and super resolution," Ph.D. Thesis, Department of Engineering Science, Univ. Oxford, 2001.
- [6] A. Criminisi, I. Reid, and A. Zisserman, "A plane measuring device," *Image and Vision Computing*, 17, 1999.
- [7] J.F. Canny. "A computational approach to edge detection," *IEEE PAMI*, Vol 8(6), 1986.
- [8] Falls Church, *Manual of Photogrammetry*, American Society of Photogrammetry, 1980.
- [9] P. Firoozfam and S. Negahdaripour, "Application of panoramic images of the seafloor for improved motion estimation and photo-mosaicking," *Proc. MTS/IEEE Oceans02*, 2002.
- [10] P. Firoozfam and S. Negahdaripour, "A conical panoramic stereo imaging system for 3-d scene reconstruction, *Proc. MTS/IEEE Oceans03*, 2003.
- [11] P. Firoozfam and S. Negahdaripour, "A multi-camera conical imaging system for robust 3-d motion estimation, positioning and mapping from UAVs," *Proc. IEEE AVSS*, Miami, FL, 2003.
- [12] P. Firoozfam, *Multi-camera imaging for 3-D mapping and positioning; Stereo and panoramic conical views*, PhD thesis, Department Electrical & Comp. Engineering, University of Miami, 2004.
- [13] M. Fischler, and R. Bolles. "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography." SRI-TN, 1980.

- [14] C.J. Harris and M. Stephens, "A combined corner and edge detector," *Proc. 4th Alvey Vision Conference*, Manchester, England, 1988.
- [15] R.I. Hartley and P. Sturm, "Triangulation. *Proc. ARPA '94*, 1994.
- [16] Hartley, R.I., "In defense of the eight point algorithm," *IEEE PAMI*, Vol 19(6), October, 1997.
- [17] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, Cambridge Press, 2000.
- [18] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," *Proc. CVPR97*, 1997.
- [19] B.K.P. Horn, *Robot Vision*, The MIT Press, 1986.
- [20] B.K.P. Horn, "Closed form solutions of absolute orientation using unit quaternions," *JOSA-A*, Vol 4(4), April 1987.
- [21] B.K.P. Horn, M. Hilden, and S. Negahdaripour, "Closed form solutions of absolute orientation using orthonormal matrices," *JOSA-A*, Vol 5(7), 1987.
- [22] K. R. Koch, *Parameter Estimation and Hypothesis Testing in Linear Models*, Springer-Verlag, 1999.
- [23] A. Leick, *GPS Satellite Surveying*, John Wiley & Sons, 1990.
- [24] H.C. Longuet-Higgins, "An algorithm for reconstructing a scene from two projections," *Nature*, 293, September, 1981.
- [25] H.C. Longuet-Higgins "The Visual Ambiguity of a Moving Plane ," *Proceedings of the Royal Society of London*, Series B, Vol. 223, 1984.
- [26] K. Levenberg. "A method for the solution of certain non-linear problems in least squares," *Quarterly Applied Mathematics*, II(2), 1944.
- [27] V. Nalwa, "A true omnidirectional viewer," AT&T Technical Report Holmdel, Bell Laboratories, 1996.
- [28] S. Negahdaripour, H. Zhang, P. Firoozfam, and J. Oles, "Utilizing panoramic views for visually guided tasks in underwater robotics applications," *Proc. MTS/IEEE Oceans*, 2001.
- [29] J. Neumann, C. Fermuller, and Y. Aloimonos, "Eyes from eyes: Analysis of camera design using plenoptic video geometry. *UMD*, 2001.

- [30] U.A. Rauhala, "Introduction to array algebra," *Photogrammetric Engineering and Remote Sensing*, Vol 46(2), February, 1980.
- [31] H.S. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment," *Proc. European Conf. Computer Vision*, 1998.
- [32] R. Swaminathan and S.K. Nayar, "Nonmetric calibration of wide-angle lenses and polycameras," *Proc. CVPR'99*, 1999.
- [33] K.H. Tan, H. Hua, and N. Ahuja, "Multiview panoramic cameras using a mirror pyramid," *Proc. OMNIVIS'02*, 2002.
- [34] P.H.S. Torr, "A Structure and Motion Toolkit in Matlab- Interactive Adventures in S and M," Tech Report MSR-TR-2002-56, Microsoft Research, Cambridge, UK, June, 2002 (also P. Torr and D. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *Int. J. Comput. Vision*, 24, 1997).
- [35] R.Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Trns. Rob. Automation*, Vol 3(4), 1987.
- [36] P. Vanicek and E. Krakiwsky. *Geodesy; The Concepts*, Elsevier Science Publishers B.V., 1986.
- [37] Z.Y. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," *Proc. ICCV99*, 1999.
- [38] <http://www.ptgrey.com/>

Appendix 1: Theoretical Estimation of Uncertainty Bounds

We present here the theoretical results on the estimation of uncertainty bounds to establish accuracy in 3-D reconstruction, motion estimation and positioning.

1.1 3-D Reconstruction by Triangulation

Given the corresponding points on two images and the camera projection matrices⁷, the point in 3-D world can be located by triangulation. Let $X = \rho[X, Y, Z, 1]^T$, $x_1 = \mu_1[x_1, y_1, 1]$, and $x_2 = \mu_2[x_2, y_2, 1]$ be a 3-D point, and its projections onto cameras 1 and 2, respectively, based on the camera projection matrices \mathcal{P}^1 and \mathcal{P}^2 :

$$\begin{cases} x_1 = \mathcal{P}^1 X \\ x_2 = \mathcal{P}^2 X \end{cases} \quad (1)$$

From many methods to solve the above equations for X [15], a close-form solution is most common. First, \mathcal{P} in (1) is combined into the form $AX = 0$, which is linear in X . Eliminating the homogenous scale factor in \mathcal{P} by a cross product gives three equations for each image point, of which two are linearly independent. For example, for the first image, $x_1 \times (\mathcal{P}^1 X) = 0$ gives

$$\begin{cases} x_1 \mathcal{P}_3^{1T} X - \mathcal{P}_1^{1T} X & = 0 \\ y_1 \mathcal{P}_3^{1T} X - \mathcal{P}_2^{1T} X & = 0 \\ x_1 \mathcal{P}_2^{1T} X - y_1 \mathcal{P}_1^{1T} X & = 0 \end{cases} \quad (2)$$

where \mathcal{P}_i^{1T} are the rows of \mathcal{P}^1 . Rewriting the above equations in the form $AX = 0$, where

$$A = \begin{bmatrix} x_1 \mathcal{P}_3^{1T} - \mathcal{P}_1^{1T} \\ y_1 \mathcal{P}_3^{1T} - \mathcal{P}_2^{1T} \\ x_2 \mathcal{P}_3^{2T} - \mathcal{P}_1^{2T} \\ y_2 \mathcal{P}_3^{2T} - \mathcal{P}_2^{2T} \end{bmatrix} \quad (3)$$

we obtain the desired solution in the form $[\rho X \ \rho Y \ \rho Z \ \rho]^T$ from the singular value decomposition of A .

With a cluster of N cameras, we have multiple observations $x_i = \mu_i[x_i, y_i, 1]$, $i = 1, 2, \dots, N$, and matrix A can be rewritten as

⁷A camera projection matrix can be readily determined through a priori calibration experiments.

$$A = \begin{bmatrix} x_1 \mathcal{P}_3^{1\top} - \mathcal{P}_1^{1\top} \\ y_1 \mathcal{P}_3^{1\top} - \mathcal{P}_2^{1\top} \\ x_2 \mathcal{P}_3^{2\top} - \mathcal{P}_1^{2\top} \\ y_2 \mathcal{P}_3^{2\top} - \mathcal{P}_2^{2\top} \\ \vdots \\ x_N \mathcal{P}_3^{N\top} - \mathcal{P}_1^{N\top} \\ y_N \mathcal{P}_3^{N\top} - \mathcal{P}_2^{N\top} \end{bmatrix} \quad (4)$$

Let $W = (A^T A) = V D V^T$ denotes the eigenvalue decomposition of the 4×4 matrix W , where V is a 4×4 orthogonal matrix, and $D = \text{diag}\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ is the diagonal matrix of eigenvalues in descending order. It can be readily shown that the 3-D point X can be determine from the eigenvector $v_4 = [v_1, v_2, v_3, v_4]^T$ corresponding to the smallest eigenvalue λ_4 .

1.1.1 Accuracy in 3-D Reconstruction

We have shown that we can estimate the position of a 3-D space point from multiple observations in several cameras. However, we are interested to know of the accuracy of the solution, given noisy measurements and the finite precision due to the spatial quantization of the image at a particular resolution. The answer can be found from the covariance matrix C_X of the estimated 3-D point X , based on the errors in the measurements $x_i = \mu_i[x_i, y_i, 1]$, $i = 1, 2, \dots, N$.

If we define the measurement vector $v = [x_1, y_1, x_2, y_2, \dots, x_N, y_N]^T$, with σ_{x_i} and σ_{y_i} , $i = 1, 2, \dots, N$, as the variances of these measurements, the covariance C_a of these elements are given by [12]:

$$C_a|_{16 \times 16} = \frac{\partial A}{\partial v} C_v \frac{\partial A^T}{\partial v} \quad (5)$$

where

$$\frac{\partial A}{\partial v} = \begin{bmatrix} \frac{\partial a_{1,1}}{\partial x_1} & \frac{\partial a_{1,1}}{\partial y_1} & \dots & \frac{\partial a_{1,1}}{\partial x_N} & \frac{\partial a_{1,1}}{\partial y_N} \\ \frac{\partial a_{1,2}}{\partial x_1} & \frac{\partial a_{1,2}}{\partial y_1} & \dots & \frac{\partial a_{1,2}}{\partial x_N} & \frac{\partial a_{1,2}}{\partial y_N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial a_{4,4}}{\partial x_1} & \frac{\partial a_{4,4}}{\partial y_1} & \dots & \frac{\partial a_{4,4}}{\partial x_N} & \frac{\partial a_{4,4}}{\partial y_N} \end{bmatrix}_{16 \times 2N} \quad (6)$$

and

$$\begin{aligned} \frac{\partial a_{m,n}}{\partial x_i} &= 2x_i \mathcal{P}_{i,3,m} \mathcal{P}_{i,3,n} - (\mathcal{P}_{i,3,m} \mathcal{P}_{i,1,n} + \mathcal{P}_{i,3,n} \mathcal{P}_{i,1,m}) \\ \frac{\partial a_{m,n}}{\partial y_i} &= 2y_i \mathcal{P}_{i,3,m} \mathcal{P}_{i,3,n} - (\mathcal{P}_{i,3,m} \mathcal{P}_{i,2,n} + \mathcal{P}_{i,3,n} \mathcal{P}_{i,2,m}) \end{aligned} \quad (7)$$

Next, it can be shown that the covariance of the reconstructed point $[\rho X, \rho Y, \rho Z, \rho]^T$ (in projective space) is

$$C_X = (V\Psi_4 V^\top H_{v_4}) C_A (V\Psi_4 V^\top H_{v_4})^\top \quad (8)$$

where

$$\Psi_4 = \text{diag}\{(\lambda_4 - \lambda_1)^{-1}, (\lambda_4 - \lambda_2)^{-1}, (\lambda_4 - \lambda_3)^{-1}, 0\} \quad (9)$$

$$H_{v_4} = \begin{bmatrix} v_1 I_{4 \times 4} & v_2 I_{4 \times 4} & v_3 I_{4 \times 4} & v_4 I_{4 \times 4} \end{bmatrix}_{4 \times 16} \quad (10)$$

where $I_{4 \times 4}$ is a 4×4 identity matrix.

Finally, the covariance of the Euclidean point $[X, Y, Z]^\top$ is $C_X = J_P C_X J_P^\top$, where J_P is the Jacobian of the transformation from projective to Euclidean space:

$$J_P = \begin{bmatrix} \frac{\partial X}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{1}{\rho} & 0 & 0 & -\frac{X}{\rho^2} \\ 0 & \frac{1}{\rho} & 0 & -\frac{Y}{\rho^2} \\ 0 & 0 & \frac{1}{\rho} & -\frac{Z}{\rho^2} \end{bmatrix}_{3 \times 4} \quad (11)$$

These analytical results allows us to estimate the reconstruction accuracy of any 3-D point on the terrain X , given its projections x_i ($i = 1, 2, \dots, N$) into the N images of the camera cluster. Arbitrary arrangement of the N -camera cluster, camera resolution, and terrain altitude are encoded in the projection matrix P_i of the i -th camera, and thus can be simulated.

1.2 3-D Motion Estimation by Absolute Orientation

We have already discussed the reconstruction of a point in 3-D space from two or more image observations. Now, suppose we know the coordinates of some 3-D points $X_{1,i}$ ($i = 1, 2, \dots, L$) at one position of the panoramic imaging system, move to a new location, and determine the position $X_{2,i}$ of the same points in the new camera coordinate system. Absolute orientation is the problem of using the measurements $X_{1,i} = [X_{1,i}, Y_{1,i}, Z_{1,i}]^\top$ and $X_{2,i} = [X_{2,i}, Y_{2,i}, Z_{2,i}]^\top$ of three or more points ($i = 1, 2, \dots, L \geq 3$) to determine the motion of the imaging system between the two views. This problem has a closed-form solution [3, 20, 21].

Let S_1 and S_2 be the two sets of measured 3-D points:

$$S_1 = \begin{bmatrix} X_{1,1} & Y_{1,1} & Z_{1,1} \\ X_{1,2} & Y_{1,2} & Z_{1,2} \\ \vdots & \vdots & \vdots \\ X_{1,L} & Y_{1,L} & Z_{1,L} \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 & Z_1 \end{bmatrix} \quad (12)$$

$$S_2 = \begin{bmatrix} X_{2,1} & Y_{2,1} & Z_{2,1} \\ X_{2,2} & Y_{2,2} & Z_{2,2} \\ \vdots & \vdots & \vdots \\ X_{2,L} & Y_{2,L} & Z_{2,L} \end{bmatrix} = \begin{bmatrix} X_2 & Y_2 & Z_2 \end{bmatrix}$$

We are looking for a transformation from the coordinate system 1 to coordinate system 2 in the form

$$\begin{bmatrix} X_{2,i} \\ Y_{2,i} \\ Z_{2,i} \end{bmatrix} = R \begin{bmatrix} X_{1,i} \\ Y_{1,i} \\ Z_{1,i} \end{bmatrix} + \mathbf{T} \quad i = 1, 2, \dots, N \quad (13)$$

where R and \mathbf{T} describes the motion of the imaging system in terms of a 3×3 rotation matrix and 3×1 translation vector. These unknowns can be computed through a *least mean square error* optimization over all the points:

$$\arg \min_{R, \mathbf{T}} \left\{ \sum_{i=1}^L \left\| \begin{bmatrix} X_{2,i} & Y_{2,i} & Z_{2,i} \end{bmatrix}^T - \left(R \begin{bmatrix} X_{1,i} & Y_{1,i} & Z_{1,i} \end{bmatrix}^T + \mathbf{T} \right) \right\|^2 \right\} \quad (14)$$

under the constraint $R^T R = I$.

It has been shown that the optimum orthonormal matrix R can be estimated from matrix M [3]:

$$\begin{aligned} M_{3 \times 3} &= (S_2 - \bar{S}_2)^T (S_1 - \bar{S}_1) \\ &= \begin{bmatrix} X_2 - \bar{X}_2 & Y_2 - \bar{Y}_2 & Z_2 - \bar{Z}_2 \end{bmatrix}^T \begin{bmatrix} X_1 - \bar{X}_1 & Y_1 - \bar{Y}_1 & Z_1 - \bar{Z}_1 \end{bmatrix} \\ &= \begin{bmatrix} X_2' & Y_2' & Z_2' \end{bmatrix}^T \begin{bmatrix} X_1' & Y_1' & Z_1' \end{bmatrix} \\ &= \begin{bmatrix} X_2'^T X_1' & X_2'^T Y_1' & X_2'^T Z_1' \\ Y_2'^T X_1' & Y_2'^T Y_1' & Y_2'^T Z_1' \\ Z_2'^T X_1' & Z_2'^T Y_1' & Z_2'^T Z_1' \end{bmatrix} \end{aligned} \quad (15)$$

where $\{\bar{\cdot}\}$ is the arithmetic mean of the values. If $M = UDV^T$ is the singular value decomposition of M , then R and \mathbf{T} are given by:

$$\begin{aligned} R &= \begin{cases} UV^T & \text{when } \det\{UV^T\} = +1 \\ U \text{diag}\{1, 1, -1\} V^T & \text{when } \det\{UV^T\} = -1 \end{cases} \\ \mathbf{T} &= \bar{S}_2 - R\bar{S}_1 \end{aligned} \quad (16)$$

Using Rodriguez formula, we can consequently determine from R the rotation vector ω :

$$\omega = \frac{\theta}{2 \sin \theta} \begin{bmatrix} r_{32} - r_{23} & r_{13} - r_{31} & r_{21} - r_{12} \end{bmatrix}^T \quad (17)$$

where $\theta = \cos^{-1}((\text{trace}(R) - 1)/2)$, and r_{ij} for $i = 1, 2, 3$ and $j = 1, 2, 3$ are the elements of the rotation matrix:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (18)$$

Conversely, we can determine the rotation matrix R from ω :

$$R(\omega) = \cos \theta I + (1 - \cos \theta) \hat{\omega} \hat{\omega}^T + \sin \theta [\hat{\omega}]_{\times} \quad (19)$$

where $\theta = |\omega|$ is the angle of rotation, $\hat{\omega} = \omega/|\omega|$ is the unit vector along the axis of direction ω , and $[\hat{\omega}]_{\times}$ is the skew symmetric matrix

$$[\hat{\omega}]_{\times} = \begin{bmatrix} 0 & -\hat{\omega}_3 & +\hat{\omega}_2 \\ +\hat{\omega}_3 & 0 & -\hat{\omega}_1 \\ -\hat{\omega}_2 & +\hat{\omega}_1 & 0 \end{bmatrix} \quad (20)$$

1.2.1 Accuracy in Motion Estimation

The precision in localizing selected 3-D terrain points, studied in section 1.1.1, directly affects how accurately we can compute the motion of the imaging system between two nearby positions. We can quantify the accuracy of the motion estimation from the absolute orientation solution in terms of the covariance matrices of ω and T . The derivation of these results uses the mixed model least-squared adjustment approach from the photogrammetry literature [22, 23, 36].

Let S_1 and S_2 be the two sets of measured 3-D points:

$$S_1 = \begin{bmatrix} X_{1,1} & Y_{1,1} & Z_{1,1} \\ X_{1,2} & Y_{1,2} & Z_{1,2} \\ \vdots & \vdots & \vdots \\ X_{1,N} & Y_{1,N} & Z_{1,N} \end{bmatrix} = [X_1 \ Y_1 \ Z_1] \quad (21)$$

$$S_2 = \begin{bmatrix} X_{2,1} & Y_{2,1} & Z_{2,1} \\ X_{2,2} & Y_{2,2} & Z_{2,2} \\ \vdots & \vdots & \vdots \\ X_{2,N} & Y_{2,N} & Z_{2,N} \end{bmatrix} = [X_2 \ Y_2 \ Z_2]$$

with C_{S_1, S_2} representing the $6L \times 6L$ covariance matrix of the 3-D points in the two coordinate systems (see results from section 1.1.1):

$$C_{S_1 S_2} = \begin{bmatrix} C_{S_1} & \\ & C_{S_2} \end{bmatrix}_{6N \times 6N} \quad (22)$$

Recall that the measurements $X_{1,i} = [X_{1,i}, Y_{1,i}, Z_{1,i}]^T$ and $X_{2,i} = [X_{2,i}, Y_{2,i}, Z_{2,i}]^T$ in the two coordinate systems of the imaging system are related by

$$R \begin{bmatrix} X_{1,i} \\ Y_{1,i} \\ Z_{1,i} \end{bmatrix} + T - \begin{bmatrix} X_{2,i} \\ Y_{2,i} \\ Z_{2,i} \end{bmatrix} = 0 \Rightarrow \quad (23)$$

$$f(\underbrace{[X_{1,i} \ Y_{1,i} \ Z_{1,i} \ X_{2,i} \ Y_{2,i} \ Z_{2,i}]^T}_{\mathbf{m}}, \underbrace{[\omega^T \ T^T]^T}_{\mathbf{x}}) = f(\mathbf{m}, \mathbf{x}) = 0$$

Determining the solution to $f(\mathbf{m}, \mathbf{x}) = 0$ for \mathbf{x} is in fact in the form of a *Mixed-Model Least Squares Adjustment* problem. To derive the covariance matrix of the estimated parameters, we define

$$B_i = \left. \frac{\partial f}{\partial \mathbf{m}} \right|_{\mathbf{m}^{(0)}, \mathbf{x}^{(0)}} = \begin{bmatrix} R & -I_{3 \times 3} \end{bmatrix}_{3 \times 6} \quad (24)$$

and

$$A_i = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{m}^{(0)}, \mathbf{x}^{(0)}} = \begin{bmatrix} X_{1,i} I_{3 \times 3} & Y_{1,i} I_{3 \times 3} & Z_{1,i} I_{3 \times 3} & I_{3 \times 3} \end{bmatrix}_{3 \times 12} \begin{bmatrix} J_R & \\ & I_{3 \times 3} \end{bmatrix}_{12 \times 6} \quad (25)$$

where $J_R = \left[\frac{\partial \omega}{\partial R} \right]_{3 \times 9}$ is the *Jacobian* calculated from the *Rodriguez rotation formula*. Matrices A and B can be defined by assembling the matrices A_i and B_i together:

$$A = \begin{bmatrix} A_1^\top & A_2^\top & \dots & A_N^\top \end{bmatrix}_{3N \times 6}^\top$$

$$B = \begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_N \end{bmatrix}_{3N \times 6N} \quad (26)$$

Finally, the covariance matrix of the estimated parameters ω and T is in the form

$$C_{\omega, T} = (A^\top (B C_{S_1, S_2} B^\top)^{-1} A)^{-1} \quad (27)$$

Special Case: If we assume an additive Gaussian noise with normal distribution $N(0, \sigma)$ for the 3-D points in sets S_1 and S_2 , (27) simplifies to

$$C_{\omega, T} = 2\sigma^2 (A^\top A)^{-1} \quad (28)$$

This equation shows that the covariance matrix of the final estimations depends on the configuration of the 3-D points and the noise level of their measurements, as well as the Jacobian matrix of the rotation matrix. The above equation can be simplified further under the small rotation assumption as follow:

$$C_{\omega, T} = 2\sigma^2 \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{12}^\top & C_{22} \end{array} \right] \quad (29)$$

where

$$\begin{aligned}
C_{11} &= \{((X_1^\top X_1 - N\bar{X}_1^2) + (Y_1^\top Y_1 - N\bar{Y}_1^2) + (Z_1^\top Z_1 - N\bar{Z}_1^2))I_{3 \times 3} - \\
&\quad \begin{bmatrix} X_1^\top X_1 - N\bar{X}_1^2 & X_1^\top Y_1 - N\bar{X}_1\bar{Y}_1 & X_1^\top Z_1 - N\bar{X}_1\bar{Z}_1 \\ X_1^\top Y_1 - N\bar{X}_1\bar{Y}_1 & Y_1^\top Y_1 - N\bar{Y}_1^2 & Y_1^\top Z_1 - N\bar{Y}_1\bar{Z}_1 \\ X_1^\top Z_1 - N\bar{X}_1\bar{Z}_1 & Y_1^\top Z_1 - N\bar{Y}_1\bar{Z}_1 & Z_1^\top Z_1 - N\bar{Z}_1^2 \end{bmatrix}\}^{-1} \\
C_{12} &= -\frac{1}{N}C_{11} \begin{bmatrix} 0 & -\sum Z_{1,i} & +\sum Y_{1,i} \\ +\sum Z_{1,i} & 0 & -\sum X_{1,i} \\ -\sum Y_{1,i} & +\sum X_{1,i} & 0 \end{bmatrix} \\
C_{22} &= \frac{1}{N} \left(I_{3 \times 3} + \begin{bmatrix} 0 & -\sum Z_{1,i} & +\sum Y_{1,i} \\ +\sum Z_{1,i} & 0 & -\sum X_{1,i} \\ -\sum Y_{1,i} & +\sum X_{1,i} & 0 \end{bmatrix} C_{12} \right)
\end{aligned} \tag{30}$$

Not surprising, the accuracy of the estimated motion parameters highly depends on the distribution of the 3-D points in space.

1.3 Positioning

If the images from multiple nearby positions are to be fused to construct a larger composite view, we need to know these positions of the imaging system, relative to some reference frame. Without loss of generality, let us assume that the reference frame is the very first position of the imaging system. The positions \mathbf{P}_i and \mathbf{P}_{i+1} in a global coordinate system—usually chosen as the initial trajectory position—of the vision system at time instants i and $i+1$ are related by

$$\mathbf{P}_{i+1} = \mathbf{R}_i \mathbf{P}_i + \mathbf{T}_i \tag{31}$$

1.3.1 Accuracy in 3-D Positioning

We are interested in finding the covariance matrix of \mathbf{P}_{i+1} , when the covariances of motion parameters and \mathbf{P}_i are known. Let \mathbf{v}_i be the variables of this transformation, i.e., $\mathbf{v}_i = [\boldsymbol{\omega}_i^\top \mathbf{T}_i^\top \mathbf{P}_i^\top]^\top$. The Jacobian matrix of this transformation is in the form

$$J_{\mathbf{P}_{i+1}} = \frac{\partial \mathbf{P}_{i+1}}{\partial \mathbf{v}} = \begin{bmatrix} X_i I_{3 \times 3} & Y_i I_{3 \times 3} & Z_i I_{3 \times 3} & I_{3 \times 3} & \mathbf{R}_i \end{bmatrix}_{3 \times 15} \begin{bmatrix} J_{\mathbf{R}_i} & \\ & I_{6 \times 6} \end{bmatrix}_{15 \times 9} \tag{32}$$

giving the covariance matrix of the estimated position:

$$\mathcal{C}_{\mathbf{P}_{i+1}} = J_{\mathbf{P}_{i+1}} \mathcal{C}_{\mathbf{v}_i} J_{\mathbf{P}_{i+1}}^\top = J_{\mathbf{P}_{i+1}} \begin{bmatrix} \mathcal{C}_{\boldsymbol{\omega}, \mathbf{T}_i} & \\ & \mathcal{C}_{\mathbf{P}_i} \end{bmatrix} J_{\mathbf{P}_{i+1}}^\top \tag{33}$$

with $\mathcal{C}_{\mathbf{P}_0} = \mathbf{0}_{3 \times 3}$.

Appendix 2: 3-D Reconstruction from Multiple 2-D Views

The estimation of motion from image sequences is one of the most important and highly investigated problems in vision literature⁸. While it is impossible study many methods in the limited time of this effort, given that we also have to explore the impact of a number of parameters, three relatively different approaches have been tested, each allowing us to incorporate GPS readings in the motion estimation process in ne way or another.

The oldest solution of the motion problem in vision literature may be the classical eight-point algorithm [24], which has been studied extensively [16, 17]. Here, correspondences of a minimum of 8 points in two images are used to determine both the motion of the camera, as well as the 3-D positions of these points in space. While the original formulation has led to other solutions with less number of points, say a nonlinear seven-point solution [2, 34], we have adopted this method since it provides a closed-form solution. It is described here for completeness, so we would show how we incorporate knowledge of GPS measurements, rather than determining it from the solution⁹. In the second approach, we incorporate the GPS readings directly into the motion model, and consequently estimate the sought after rotational from a non-linear estimation process. Finally, we apply a closed-form solution based on the small rotation approximation– utilizing a constraint similar to the differential optical flow model. To the best of our knowledge, a similar solution has not been proposed before.

2.1 Closed-Form Eight-Point Solution

Consider the projection $p_k = (x_k, y_k)^T$ of a scene point $P = (X, Y, Z)^T$:

$$\bar{p}_k \cong C_k \bar{P},$$

where \cong denotes up to scale equality, $\bar{p}_k \cong (p_k^T, 1)^T$ and $\bar{P} = (P^T, 1)^T$ are the homogenous coordinates of the image point p_k and scene point P , respectively (p is determined from \bar{p}_k by dividing by the 3rd component), and C_k is known as the camera matrix:

$$C_k = M_{int} K [R_{ok} | t_{ok}].$$

Here, M_{int} is the 3×3 matrix of camera intrinsic parameters¹⁰, $K = [I_{3 \times 3} | 0_{3 \times 1}]$, and $\{R_{ok} | t_{ok}\}$ describe the pose and position of the camera at position k , relative to the refer-

⁸Hartley and Zisserman [17] is the probably the best reference for many of the most recent feature-based methods.

⁹Our simulations show that we would obtain more accurate results, given a relatively large distance– in the order of 1/5-1/10 of the fight altitude– between two aircraft positions, known with variances of about 1-2 m.

¹⁰We assume a calibrate imaging system, and thus this is known.

ence coordinate system. It is well-known that p_i and p_j satisfy the epipolar constraint

$$p_j^T E_{ij} p_i = 0,$$

where $E_{ij} = [t_{ij}]_{\times} R_{ij}$ is commonly known as the *essential matrix*, and 3×3 matrix $[t]_{\times}$ represents the 3-D vector t as a 3×3 skew-symmetric matrix, such that $[t]_{\times} x = t \times x$ for any 3-D vector x . With 8 or more point correspondences $\{p_i^k, p_j^k\}$ ($k = 1 : N \geq 8$), we can write enough linear homogeneous equations in the form of the above constraint, which can subsequently be solved for the 9 unknowns in E_{ij} , up to a scale factor ambiguity¹¹. While the next stage involves determining R_{ij} and t_{ij} from E_{ij} , we show one way for incorporating the knowledge of t_{ij} to compute R_{ij} .

To do this, we first normalize E_{ij} according to $E_{ij} = \sqrt{2/\text{trace}(E_{ij}^T E_{ij})} E_{ij}$ (we have used the same notation for E_{ij} before and after scaling). With e_k ($k = 1 : 3$) denoting the rows of E_{ij} , and $\omega_k = e_k \times t_{ij}$, the rows r_i ($i = 1 : 3$) of the rotation matrix are given by $r_i = -(\omega_i + \omega_j \times \omega_k)$, where $\{i, j, k\}$ is an even permutation of $\{1, 2, 3\}$. Finally, given that R_{ij} may not be a rotation matrix due to noises of feature correspondence positions and translation vector, the best approximate rotation matrix is found as $R_{ij} = UV^T$, where U and V are the left and right matrices in the singular value decomposition of $R_{ij} = U\Lambda V^T$ (again, same notation is used before and after "singular value normalization."

2.2 Non-Linear Solution

Here, we start with the projection equation:

$$\bar{p}_k = C_k \bar{P}; \quad C_k = M_{int} K [R_{ok} | t_{ok}].$$

Each image point $\bar{p}_k = (x_k, y_k, 1)^T$ gives two constraints in terms of the 3 independent unknowns of the rotational motion R_{ok} ¹², and 3 unknowns for each terrain point $\bar{P} = (X, Y, Z, 1)^T$. Suppose we have $k = 1 : M$ views of $l = 1 : N$ points. With 3 unknowns of each view's pose angles (rotation matrix) and 3 for each 3-D terrain point, we have $3 * M + 3 * N$ unknowns with $2 * M * N$ equations. If we assume the camera coordinate system at some view m as to be the reference frame, we have $R_{om} = I_{3 \times 3}$ and are thus down to $3 * M + 3 * N - 3$ unknowns. With $M = 2$ views, we need a minimum of $N = 3$ points to have sufficient equations to solve for the 12 unknowns from 12 nonlinear equations; each in the form of the above constraint. A common approach is to apply the Levenberg-Marquardt optimization technique, providing the analytical Jacobian of the constraint equation to speed up the convergence [26].

¹¹This ambiguity will be resolved with knowledge of t_{ij} , as shown later.

¹²We can express R_{ok} in terms its 3 degrees of freedom in many ways, including the Rodriguez formula involving the axis and angle of rotation.

2.3 Closed-Form Solution with Small-Angle Rotations

Again, the projection $p_k = (x_k, y_k)^T$ of a scene point $P = (X, Y, Z)^T$ gives

$$\bar{p}_k = M_{int}K[R_{ok}|t_{ok}]\bar{P}.$$

The first view may be assumed to be at the reference position: $R_{o1} = I_{3 \times 3}$, and $t_{o1} = 0$. Thus, we can write $\bar{p}_1 = M_{int}P$, which gives two constraints on $P = (X, Y, Z)^T$. If m_i ($i = 1 : 3$) denotes rows of M_{int} , and defining 3-D vectors $\alpha_1 = x_1 m_3 - m_1$ (with components α_{1i} ; $i = 1 : 3$) and $\beta_1 = y_1 m_3 - m_2$ (with components β_{1i} ; $i = 1 : 3$), it can be readily shown that

$$X = k_x Z \quad \text{and} \quad Y = k_y Z,$$

where

$$k_x = (\alpha_{12}\beta_{13} - \alpha_{13}\beta_{12})/(\alpha_{11}\beta_{12} - \alpha_{12}\beta_{11})$$

$$k_y = (\alpha_{13}\beta_{11} - \alpha_{11}\beta_{13})/(\alpha_{11}\beta_{12} - \alpha_{12}\beta_{11}).$$

For the 2nd view, we have

$$\bar{p}_2 = M_{int}(R_{o2}P + t_{o2}).$$

If rotation is assumed small¹³, we can use the approximation $R_{o2} = I + [\omega]_\times$, where $[\omega]_\times$ is a skew-symmetric matrix corresponding to the rotation vector ω : ($[\omega]_\times x = \omega \times x$). By substitution and some tedious algebraic manipulation, we finally arrive at two constraint equations:

$$[(\alpha_{23}k_y - \alpha_{22})\omega_x + (\alpha_{21} - \alpha_{23}k_x)\omega_y + (\alpha_{22}k_x - \alpha_{21}k_y)\omega_z + (\alpha_{21}k_x + \alpha_{22}k_y + \alpha_{23})]Z + g_a = 0,$$

$$[(\beta_{23}k_y - \beta_{22})\omega_x + (\beta_{21} - \beta_{23}k_x)\omega_y + (\beta_{22}k_x - \beta_{21}k_y)\omega_z + (\beta_{21}k_x + \beta_{22}k_y + \beta_{23})]Z + g_b = 0$$

where

$$g_a = \alpha_2 \cdot t_{o2} \quad g_b = \beta_2 \cdot t_{o2},$$

$\alpha_2 = x_2 m_3 - m_1$ (with components α_{2i} ; $i = 1 : 3$), $\beta_2 = y_2 m_3 - m_2$ (with components β_{2i} ; $i = 1 : 3$), and $\omega = (\omega_x, \omega_y, \omega_z)^T$ and $t_{o2} = (t_x, t_y, t_z)^T$ are described in terms of their motion components. These equations resemble the differential image motion constraint [19, 24], written in terms of point correspondences $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$.

Eliminating Z leads to a constraint equation in terms of the rotation ω :

$$\begin{aligned} & ((\beta_{23}k_y - \beta_{22})g_a - (\alpha_{23}k_y - \alpha_{22})g_b)\omega_x + \\ & ((\beta_{21} - \beta_{23}k_x)g_a - (\alpha_{21} - \alpha_{23}k_x)g_b)\omega_y + \\ & ((\beta_{22}k_x - \beta_{21}k_y)g_a - (\alpha_{22}k_x - \alpha_{21}k_y)g_b)\omega_z + \\ & (\beta_{21}k_x + \beta_{22}k_y + \beta_{23})g_a - (\alpha_{21}k_x + \alpha_{22}k_y + \alpha_{23})g_b = 0. \end{aligned}$$

¹³In practice, the raw image can be de-rotated based on some a prior rough knowledge of pitch, roll and yaw angles, e.g., either from gyros or an earlier estimate of the angles at previous UAV position, leaving us with small correction angles to be estimated.

With GPS readings which gives us t_{o2} , this becomes a linear constraint in terms of the rotational motion components. Minimum of 3 point correspondences enable us to compute a solution in closed form. A least-square solution can be determined from a redundant set of equations with $N > 3$ points.

2.4 Construction of Terrain 3-D Points

Once we know the motion, we can write the constraint $\bar{p}_k = C_k \bar{P}$ in the form

$$\begin{aligned} (x_k c_k^3 - c_k^1) \cdot \bar{P} &= 0 \\ (y_k c_k^3 - c_k^2) \cdot \bar{P} &= 0 \end{aligned}$$

where c_k^i ($i = 1 : 3$) denote the rows of the camera matrix C_k . With $M \geq 2$ views, we have a redundant set of equations to solve for \bar{P} :

$$A\bar{P} = 0; \quad A = \begin{bmatrix} x_1 c_1^3 - c_1^1 \\ y_1 c_1^3 - c_1^2 \\ x_2 c_2^3 - c_2^1 \\ y_2 c_2^3 - c_2^2 \\ \vdots \\ x_M c_M^3 - c_M^1 \\ y_M c_M^3 - c_M^2 \end{bmatrix}_{2M \times 4}.$$

The up-to-scale solution can be found from the eigenvector corresponding to the smallest eigenvalue of the 4×4 matrix $(A^T A)$. Scale ambiguity is resolved by finding P from the first 3 elements of \bar{P}/\bar{P}_4 .

2.5 Estimation of Motion from Known Target Points

Assume we know the position P^i ($i = 1 : N$) of some targets on the ground¹⁴. We want to determine the pose (angular motions relative to reference frame) of the aircraft, while also utilizing the GPS readings. Again, we start from the projection equation:

$$\bar{p}_k^i = M_k[R_{ok}|t_{ok}]\bar{P}^i,$$

which give two constraints in terms of only 3 independent unknowns in the rotational motion R_{ok} . With M views of N points, we have $3 \cdot M - 3$ unknowns in $2 \cdot M \cdot N$ equations. Even with 3 points, we have more equations than unknowns ($6M > 3M - 3$). While the equations are nonlinear, we can again apply the Levenberg-Marquardt optimization technique, supplying the analytical Jacobian for faster convergence.

¹⁴These can be determined from the earlier solutions.

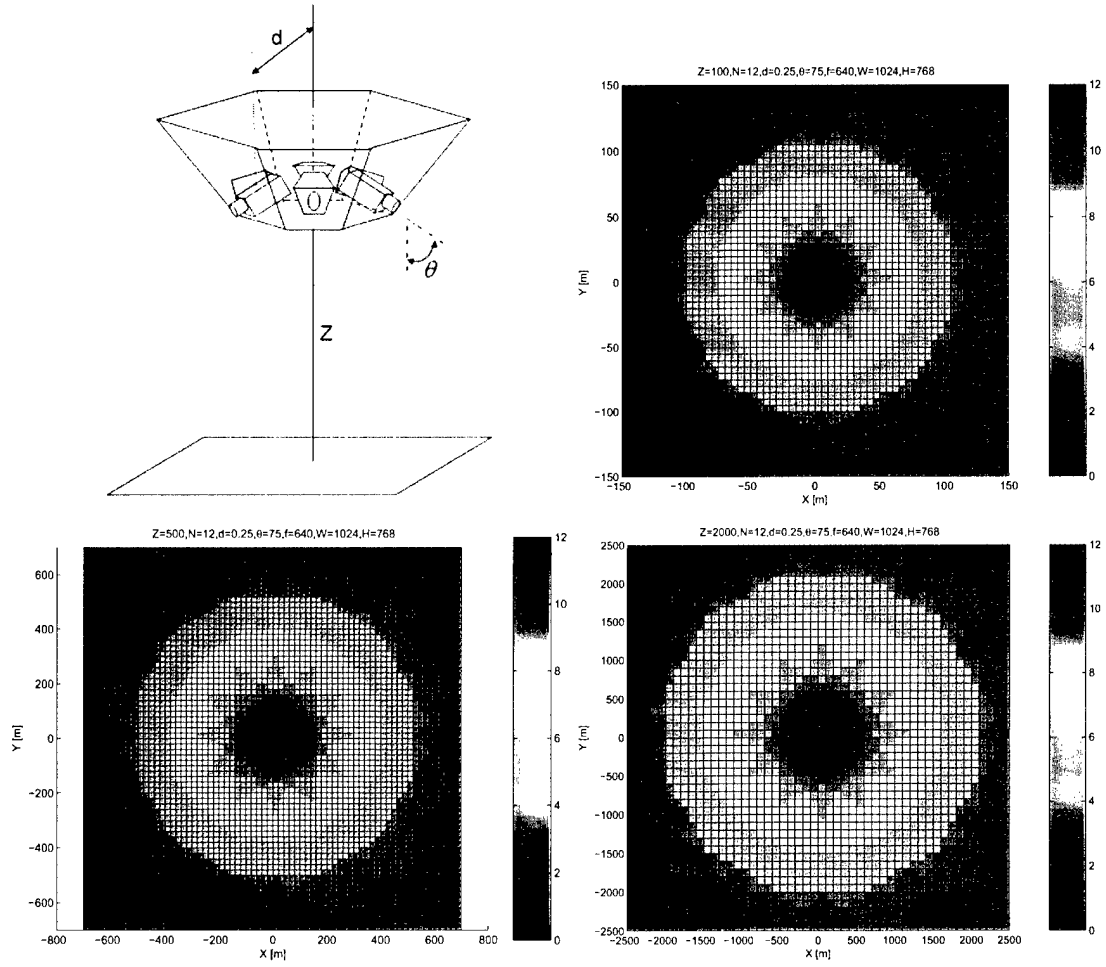


Figure 1: Schematic diagram of a down-look camera cluster for the construction of conical panoramic views, depicted with 6 cameras as an example. The three maps display the terrain regions viewed by various number of cameras in a 12-camera panoramic imaging system, at 3 different UAV elevations of 100, 500 and 2000 meters.

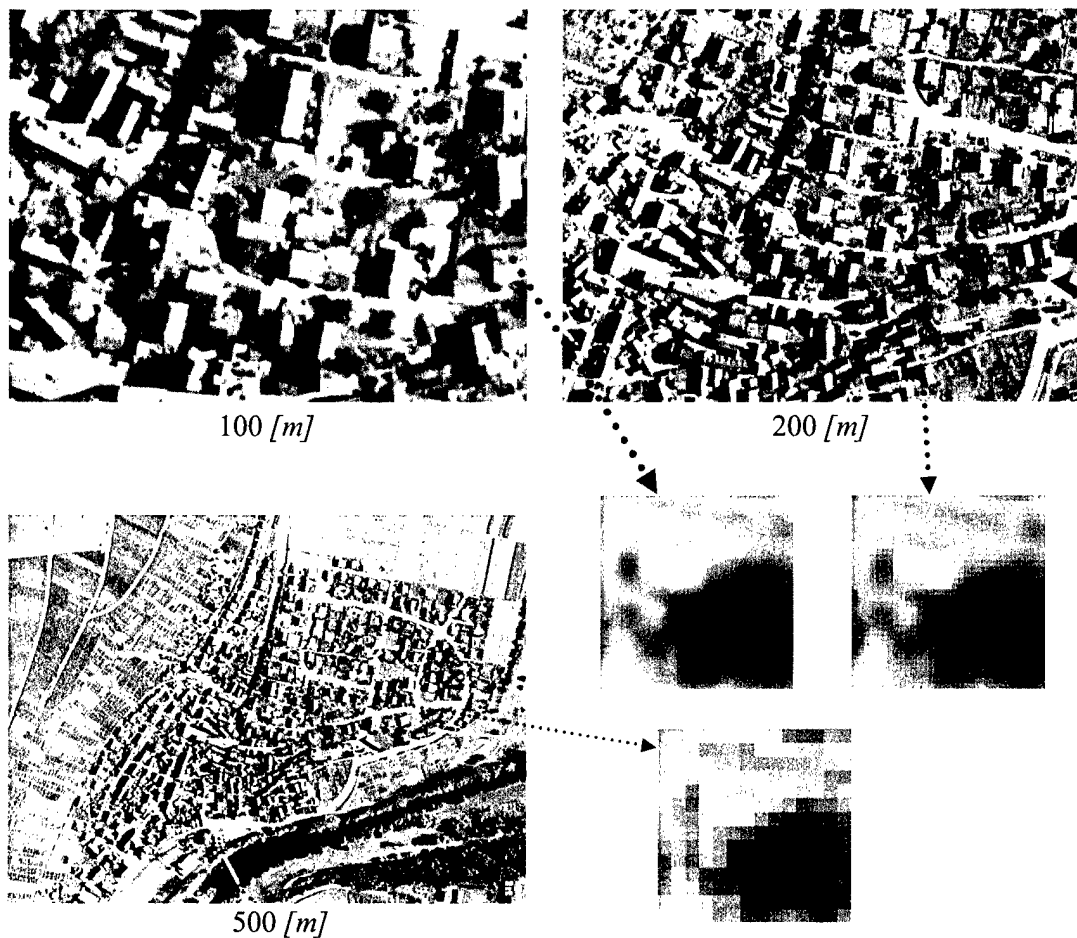


Figure 3: Images showing relative terrain coverage with fixed camera resolution and field of view at altitudes 100, 200, and and 500 meters, also comparing image resolutions over a terrain feature.

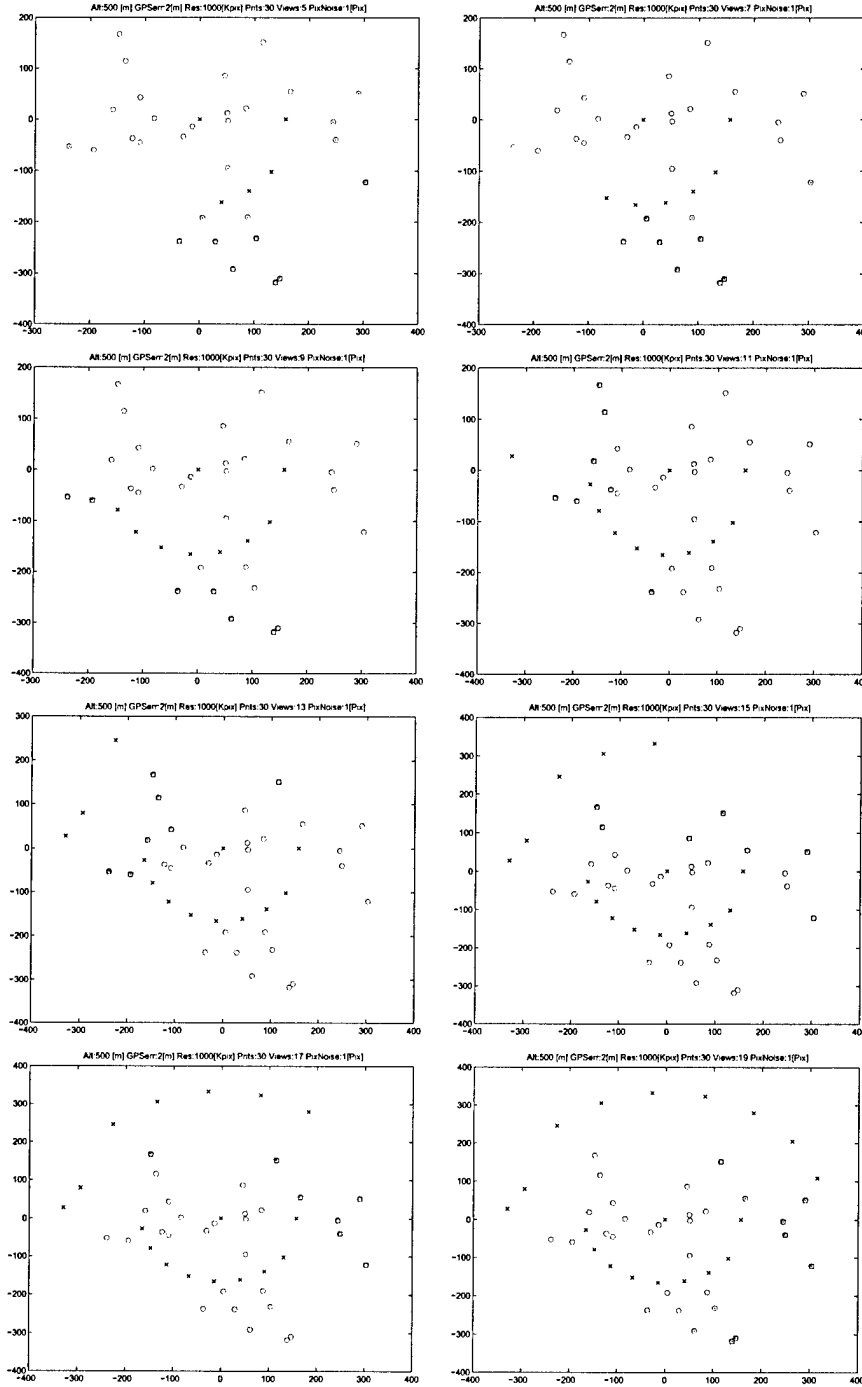


Figure 4: An experiment at 500 [m], showing XY projections of 30 ground features (red circles) imaged from 5, 7, ..., 19 UAV positions (black x). 25% features with highest error variance (blue square) and those 3 times above the median (red dot) are identified, typically located near the boundaries either far away or below a cluster of UAV viewing positions.

Appendix 3a: Simulations of panoramic high-resolution imaging system

Results in this appendix comprise simulations to test the performance of the panoramic imaging system, by varying the values of various design parameters. These tests deal with both the estimation of the 3-D coordinates of terrain features, and 3-D motion utilizing 3-D feature coordinates calculated at each UAV position along its trajectory.

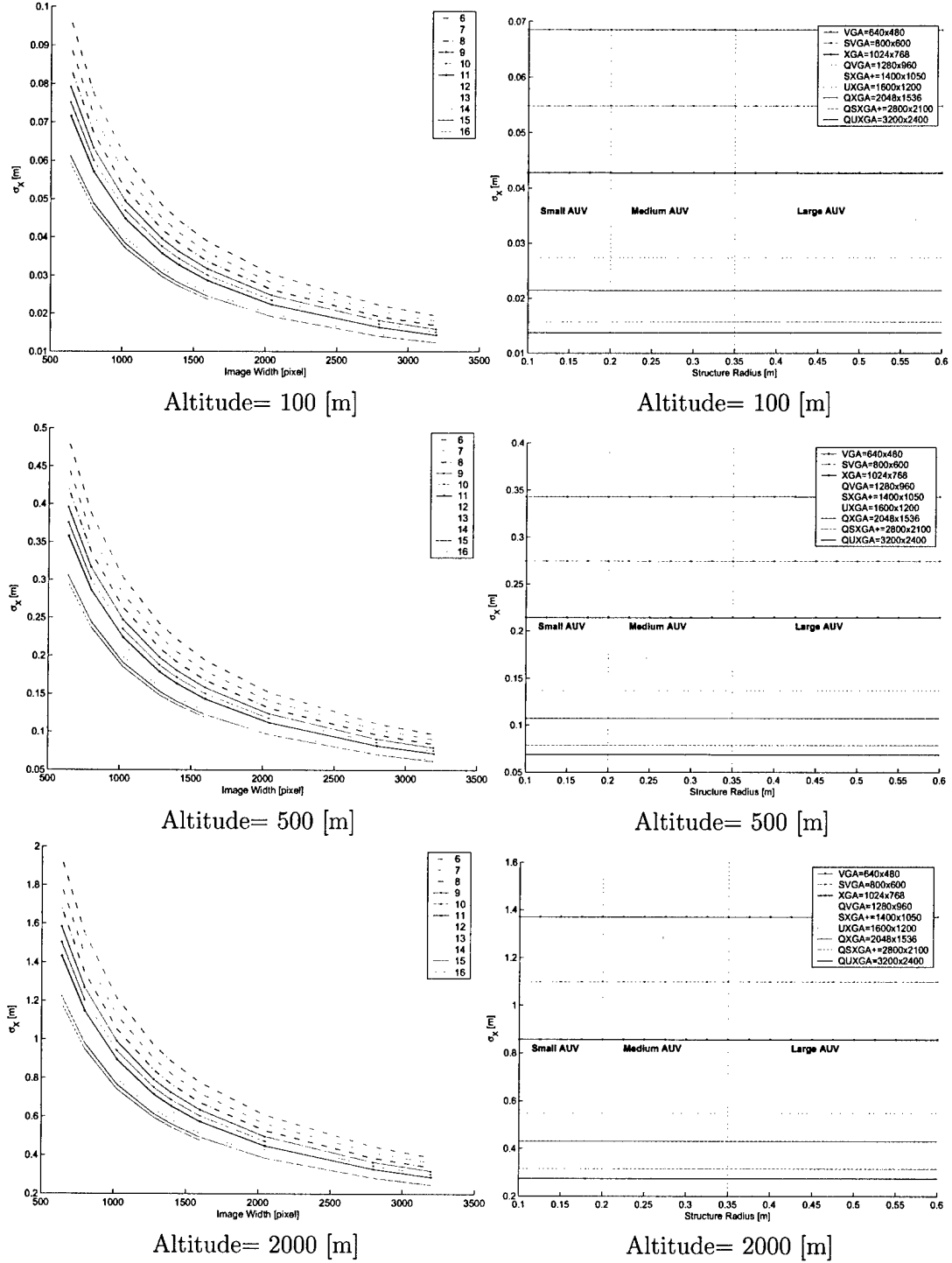


Figure 4: Uncertainty in the estimated x component of the terrain feature position at 3 UAV altitudes of 100 [m], 500 [m] and 2 [km] for various system parameters, including number of cameras, camera resolution, and imaging system radius that controls the camera baselines.

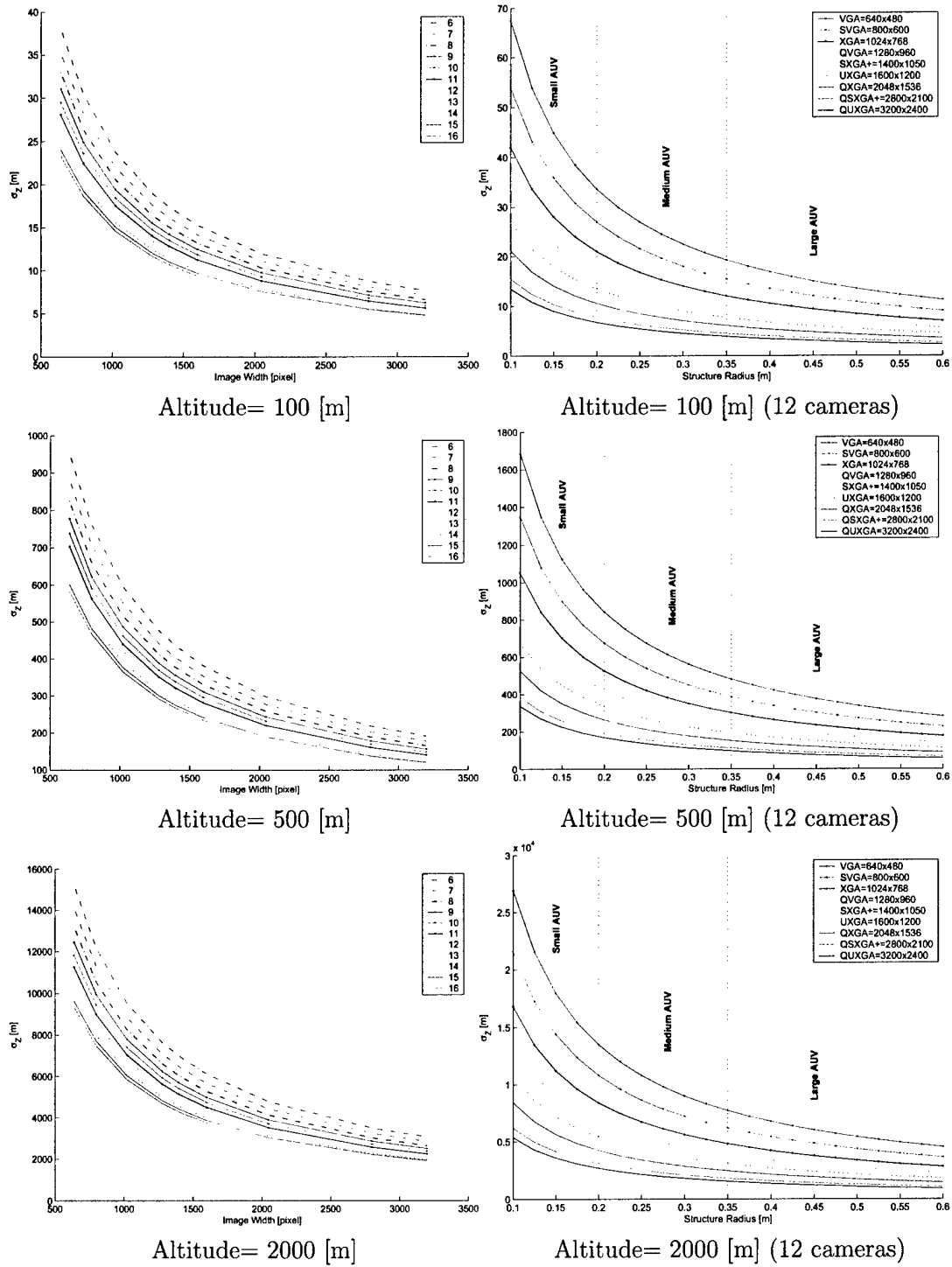


Figure 5: Uncertainty in the estimated Z component of the terrain feature position at 3 UAV altitudes of 100 [m], 500 [m] and 2 [km] for various system parameters, including number of cameras, camera resolution, and imaging system radius that controls the camera baselines.

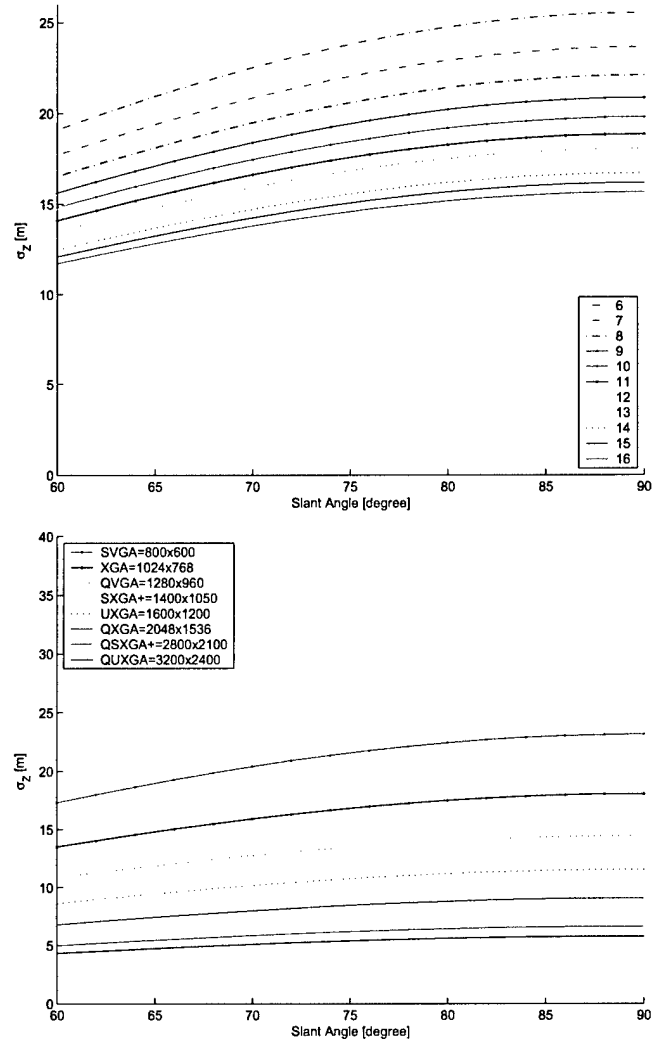


Figure 6: Variation in localization uncertainty when varying the camera slant angle θ based on the number (top) and resolution (bottom) of the cameras.

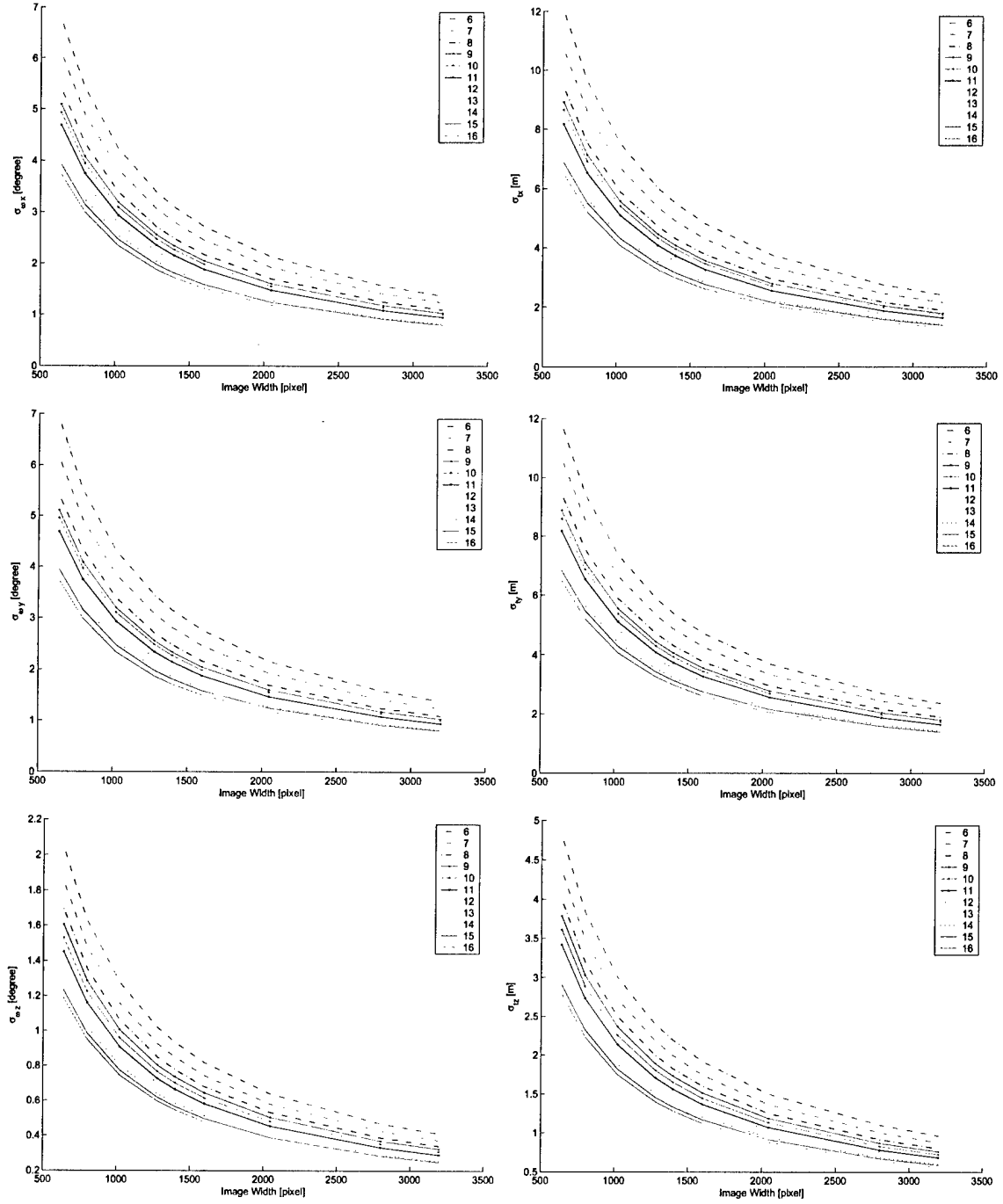


Figure 7: Uncertainties in the estimation of 3-D motion parameters, rotation $[\omega_x, \omega_y, \omega_z]$ and translation $[t_x, t_y, t_z]$, while varying camera resolution and number of cameras.

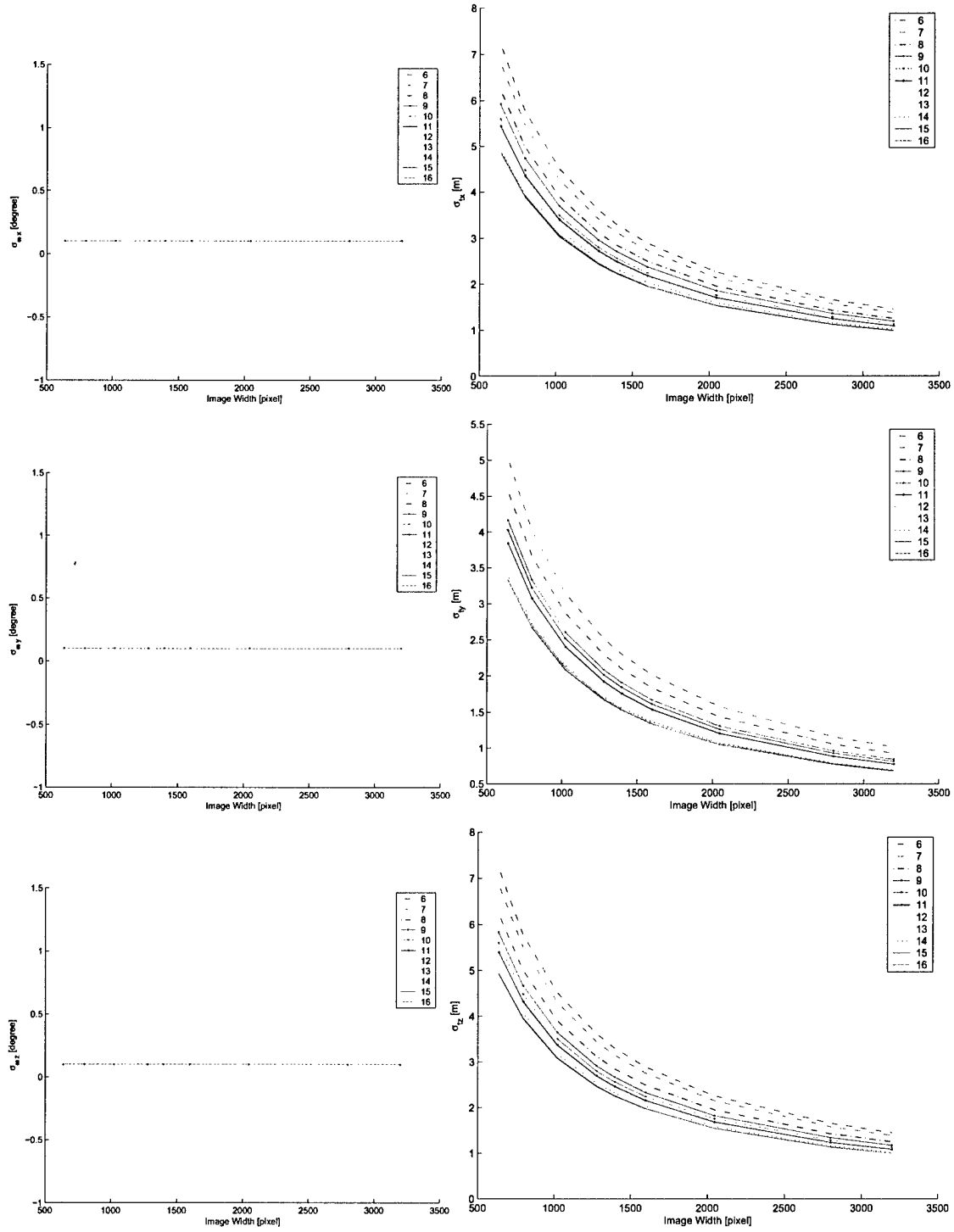


Figure 8: Uncertainties in estimating the 3-D translation components, assuming knowledge of rotational motion with an uncertainty of 0.1 [deg], while varying the number and resolution of cameras.

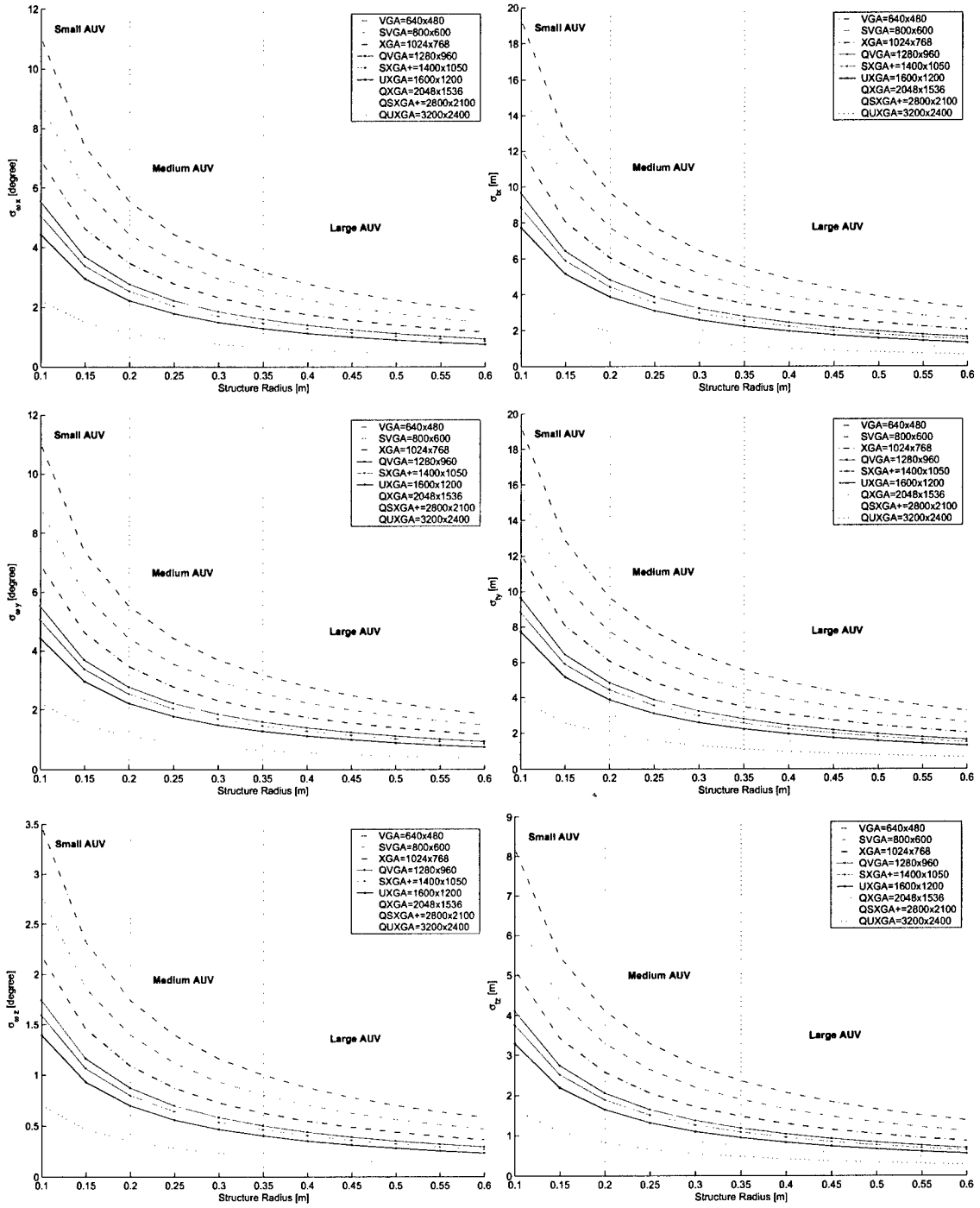


Figure 9: Uncertainties of 3-D motion parameters, rotation $[\omega_x, \omega_y, \omega_z]$ and translation $[t_x, t_y, t_z]$, varying radius of imaging system structure with 12 cameras for various camera resolutions.

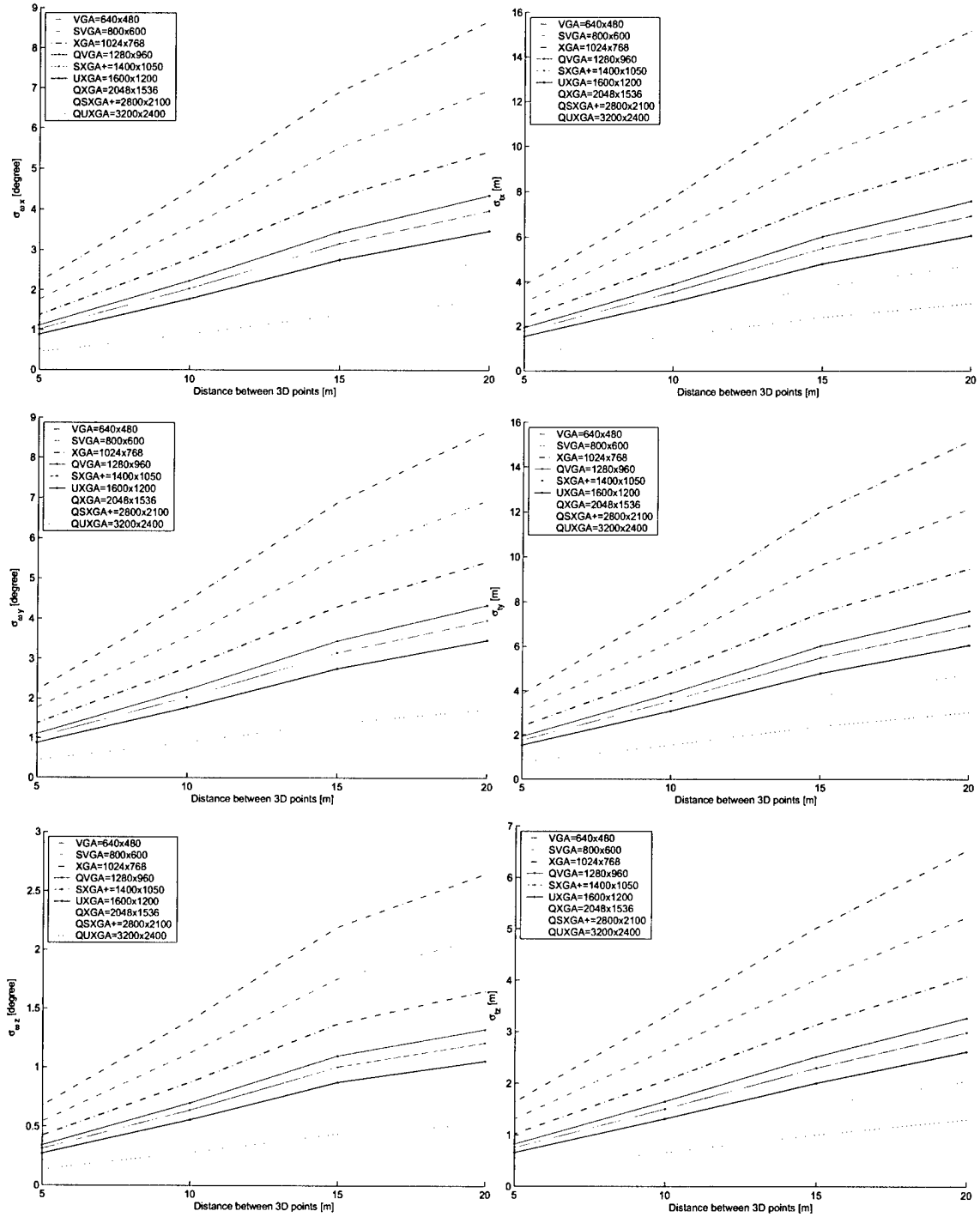


Figure 10: Uncertainties of 3-D motion parameters, rotation $[\omega_x, \omega_y, \omega_z]$ and translation $[t_x, t_y, t_z]$, varying density of terrain feature points used in the computations for various camera resolutions.

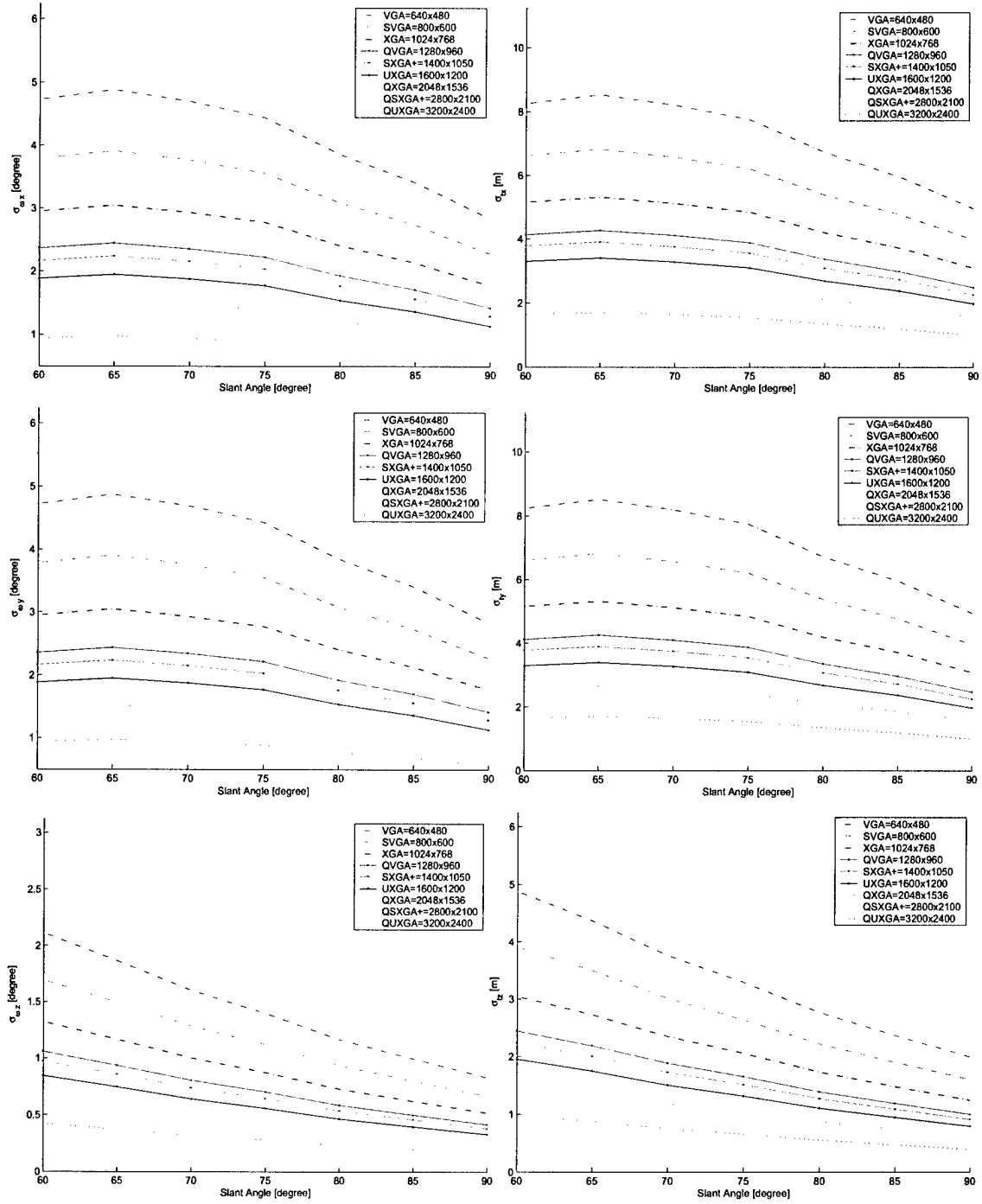


Figure 11: Uncertainties in the estimation of 3-D motion parameters, rotation $[\omega_x, \omega_y, \omega_z]$ and translation $[t_x, t_y, t_z]$, with varying the viewing angle of 12 cameras at various resolutions.

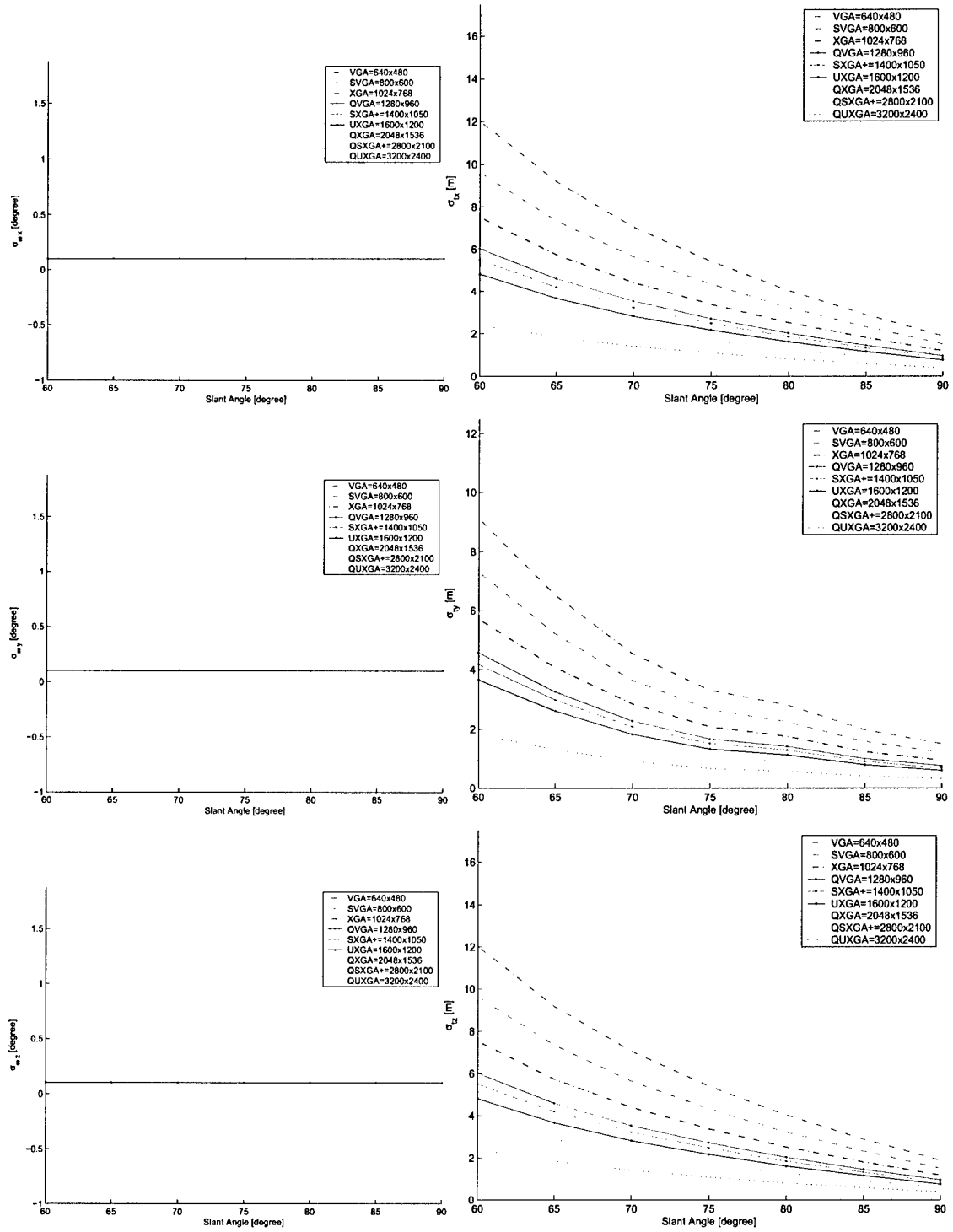


Figure 12: Assuming knowledge of rotation with an uncertainty of 0.1 [deg], 3-D translation uncertainties are determined while varying the viewing angle of 12 cameras at various resolutions.

Appendix 3b: Results of closed-form eight-Point algorithm for terrain feature localization

Results in this appendix are for the three altitudes of 500 [m], 2000 [m] and 4000 [m]. Each page comprise 5×4 arrays of plots arranged in 3 rows and 2 columns. Each row corresponds to one of the three $L \times L$ ($L = \{1, 3, 4\}$) image resolutions. Each of the two columns deal with one of N ($N = \{15, 30, 45, 60\}$) number of terrain features that are used in the computation of the UAV pose from two views. The complete set for all four choices for N are given on two subsequent pages. Each of the 5×4 arrays correspond to computations based on tracking the features in M ($M = 2 : 21$) views. Finally, various pages contain the results for GPS variances $\sigma_{GPS} = \{0, 1, 2\}$ [m].

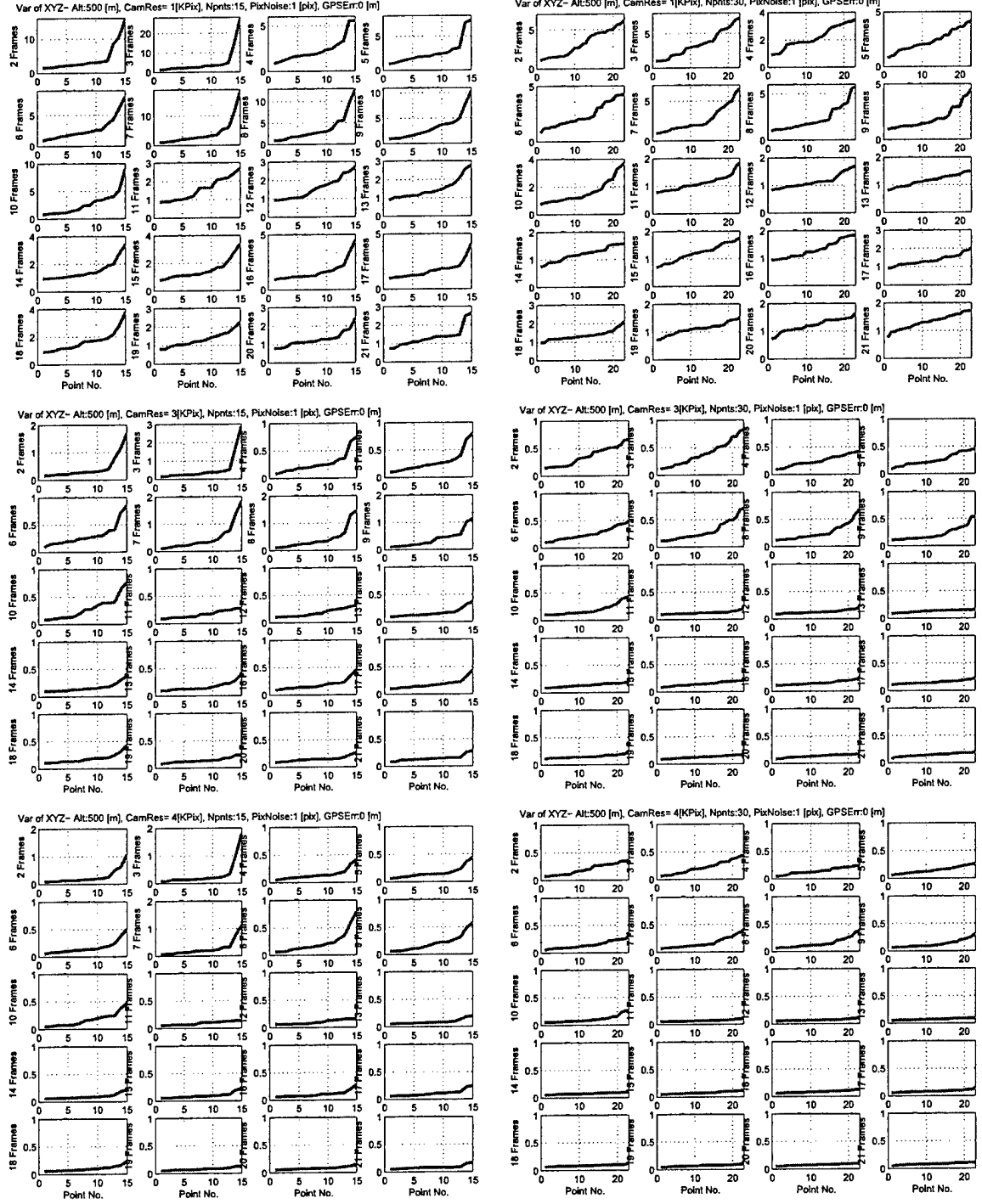


Figure 13: Closed-Form 8-Point Algorithm– Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 500); See text for details.

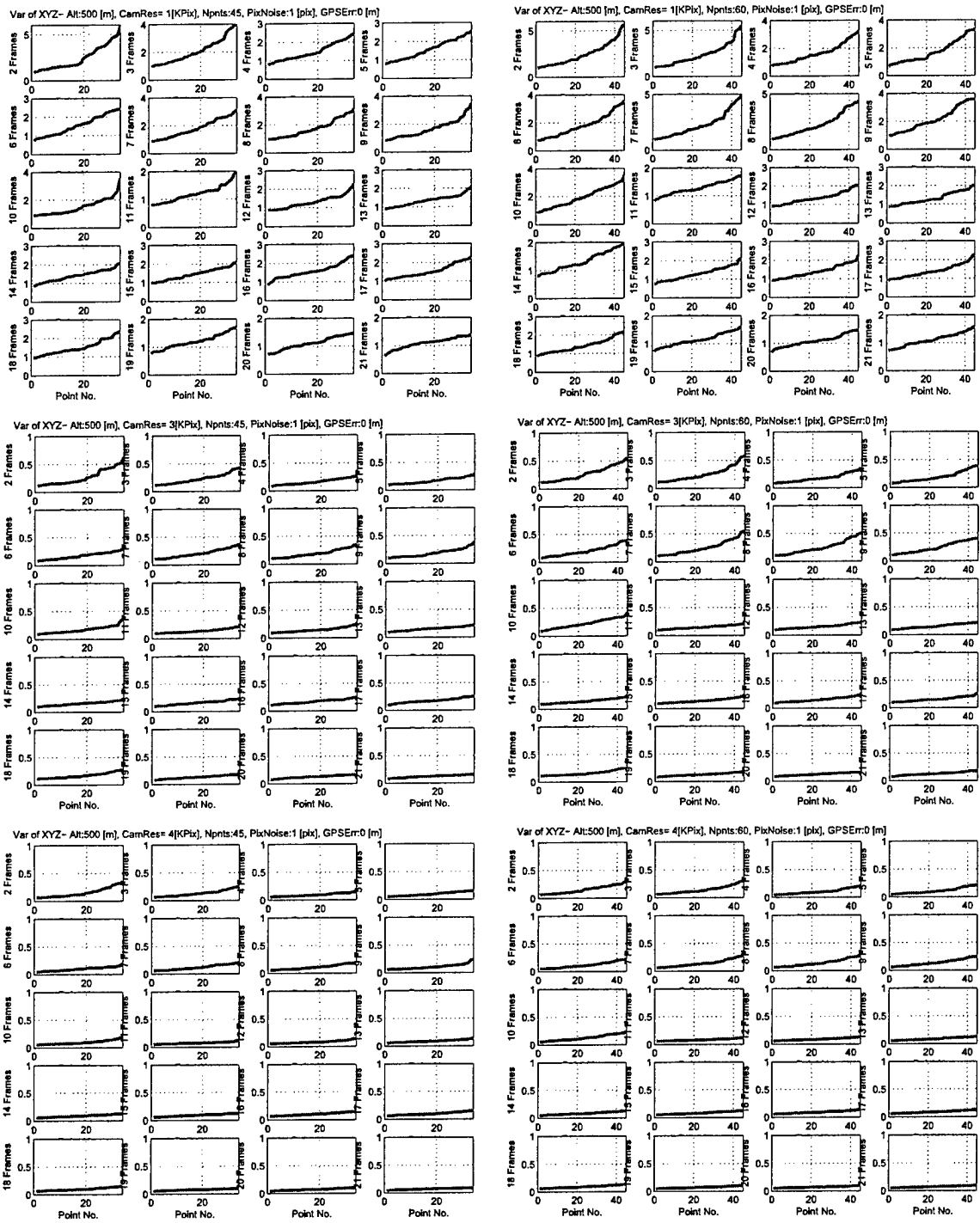


Figure 13: (continued)– Tracking 45 (left) and 60 (right) points with noise-free GPS.



Figure 13: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

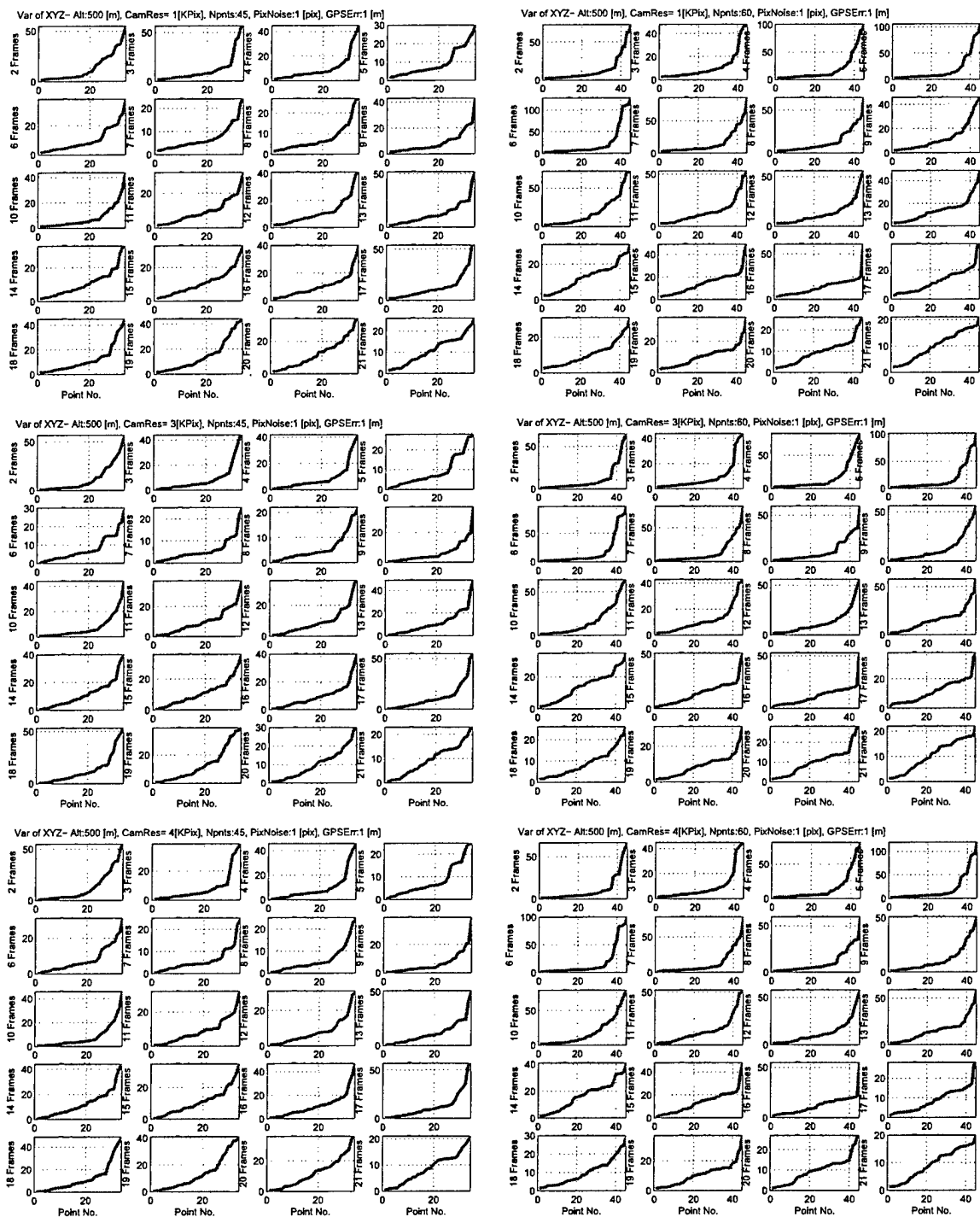


Figure 13: (continued)– Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

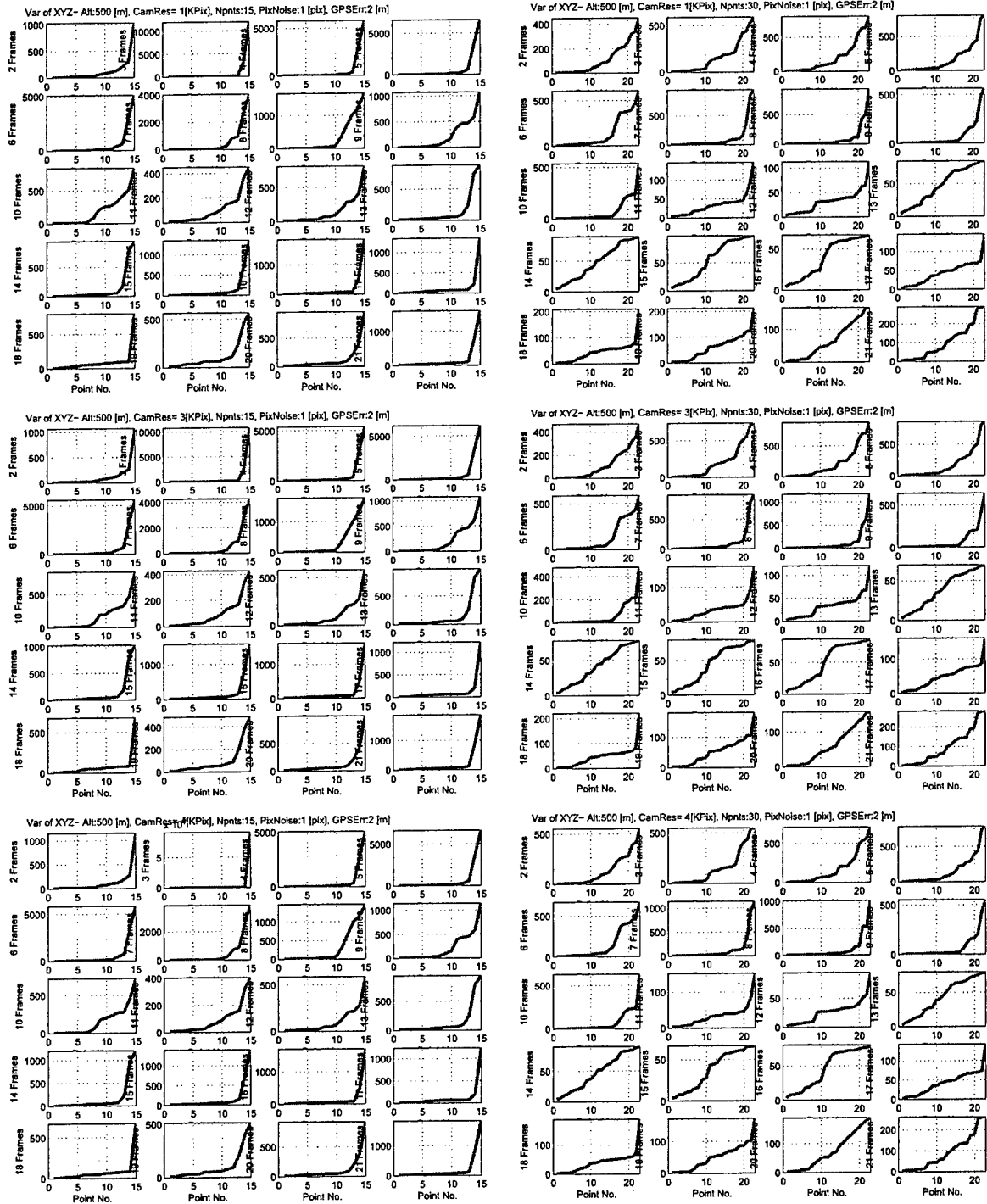


Figure 13: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

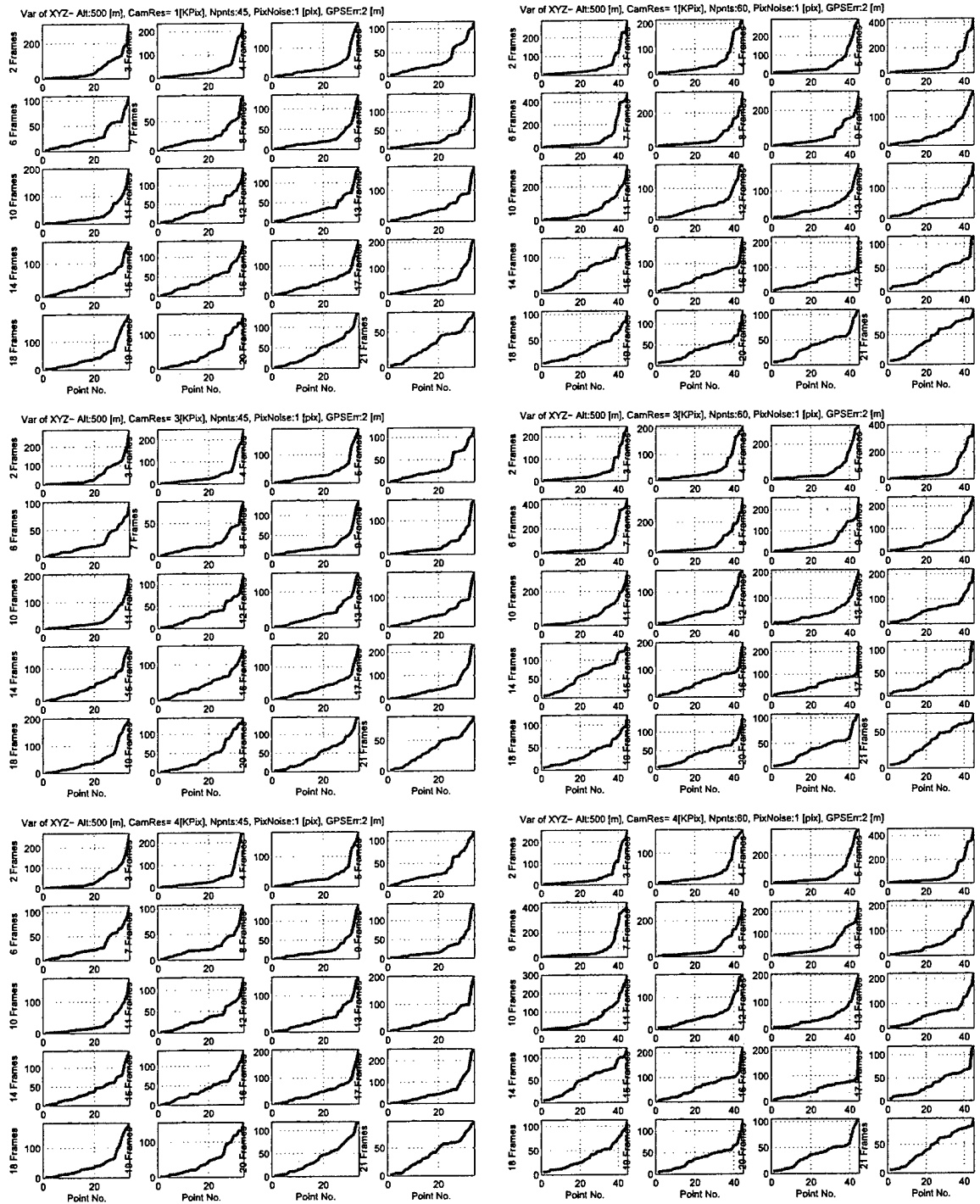


Figure 13: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

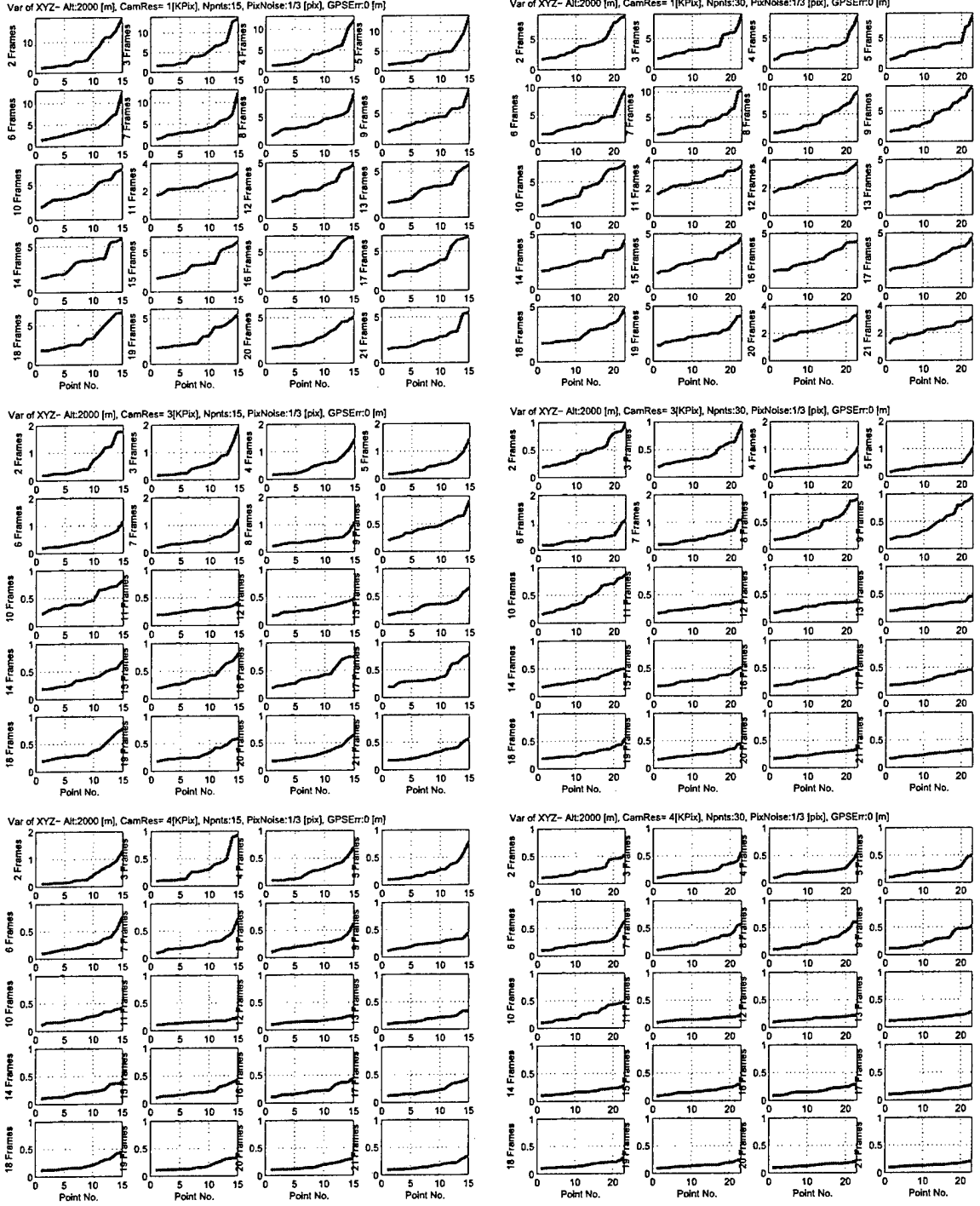


Figure 14: Closed-Form 8-Point Algorithm– Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (altitude of 2000 [m]).

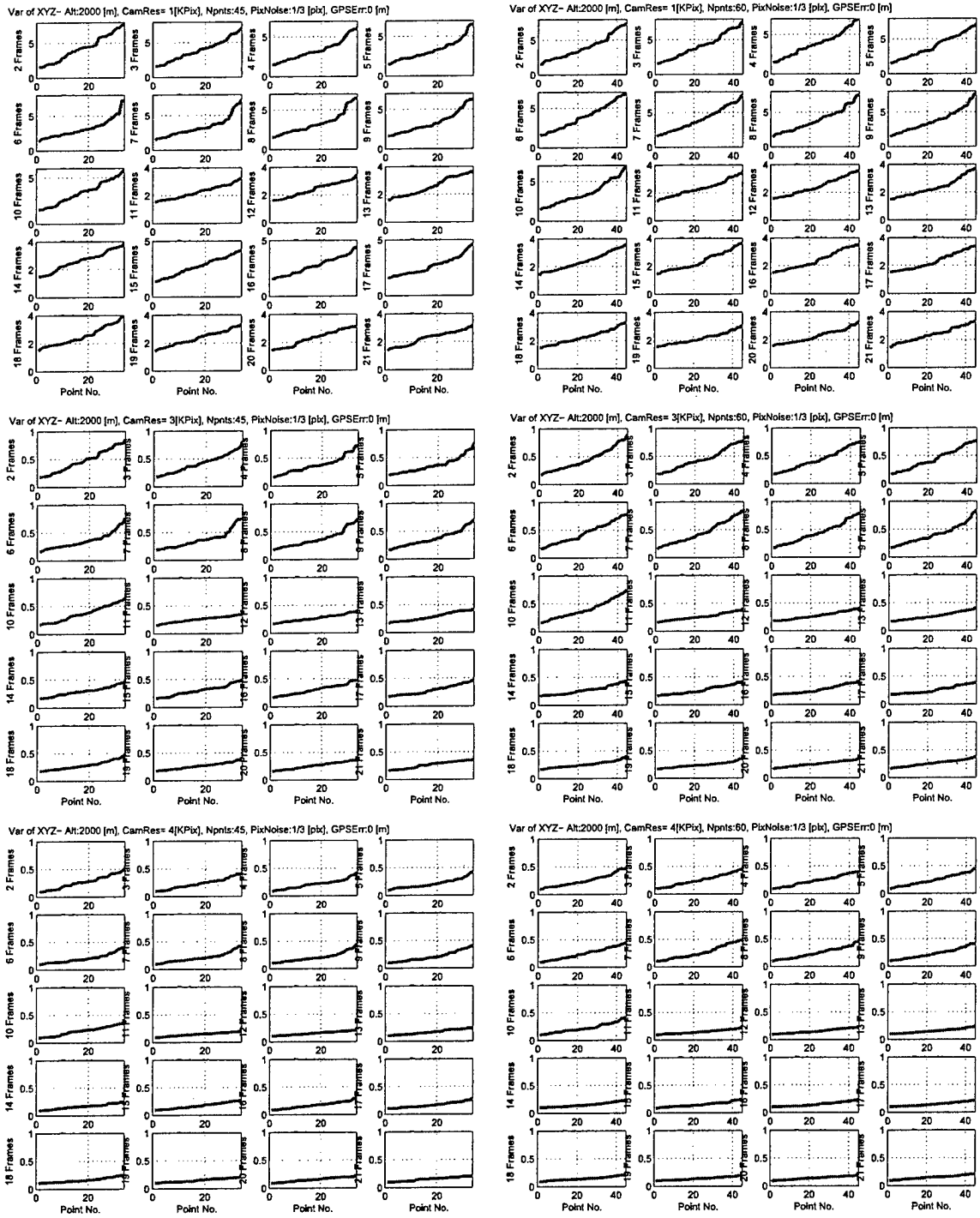


Figure 14: (continued)- Tracking 45 (left) and 60 (right) points with noise-free GPS.

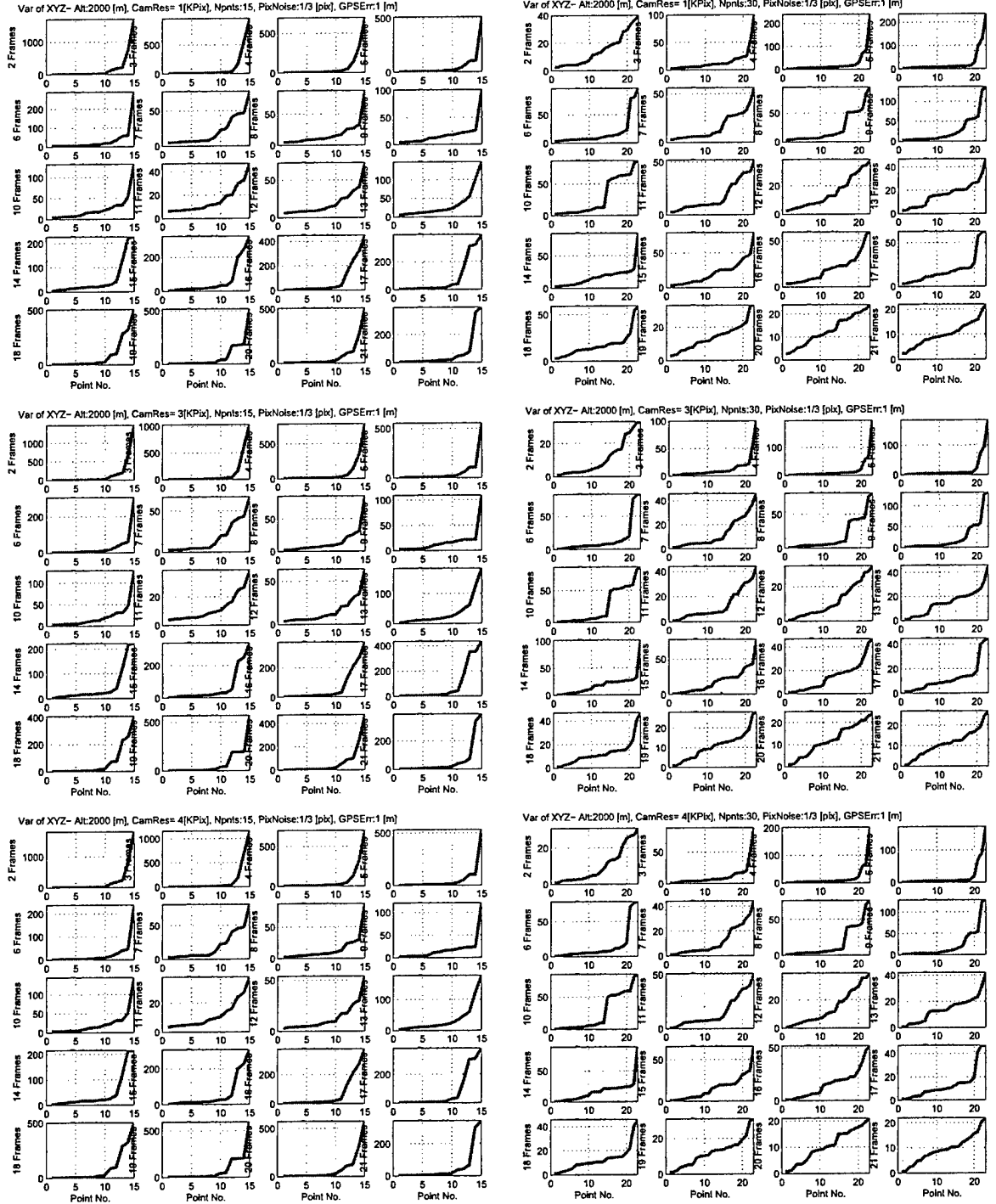


Figure 14: (continued)– Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

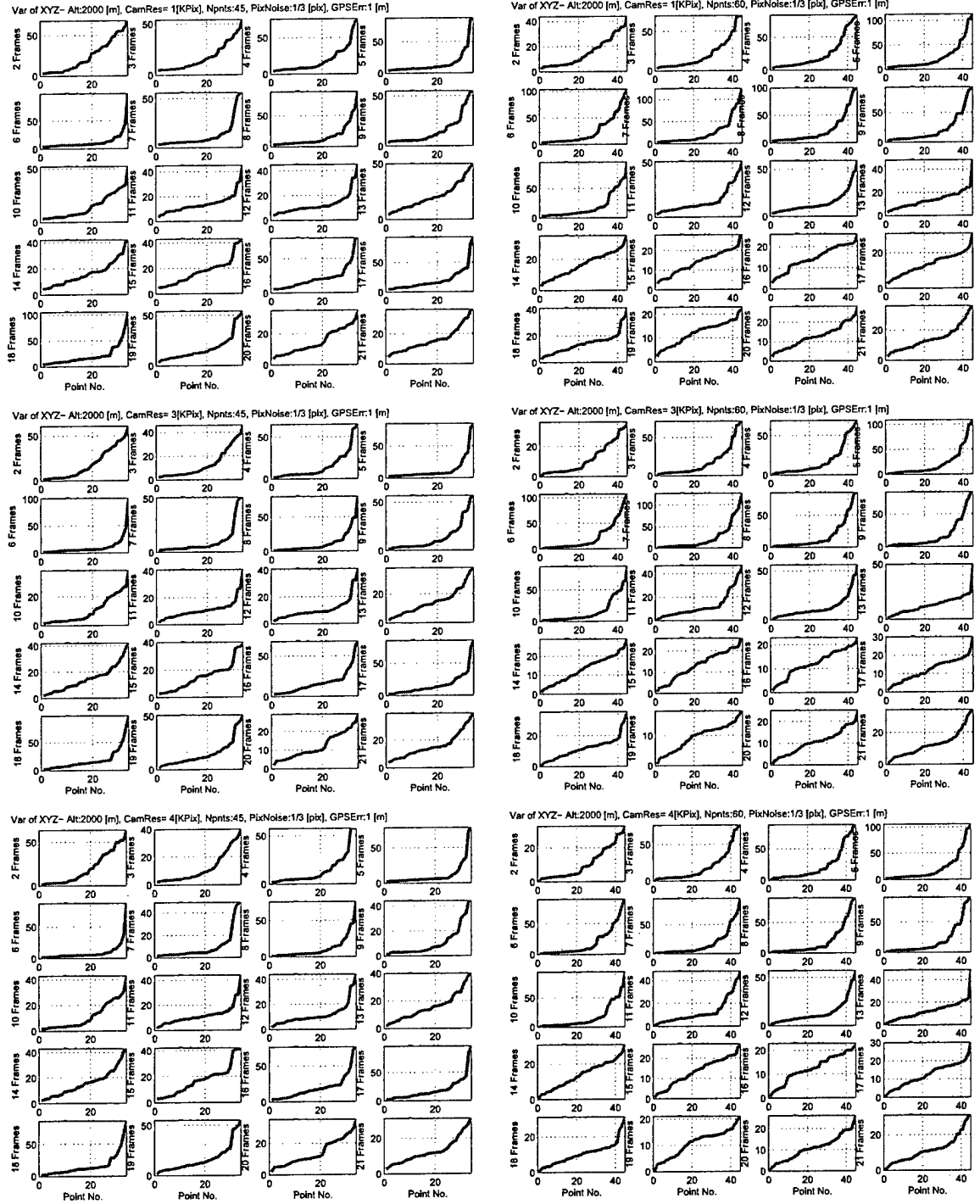


Figure 14: (continued)– Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

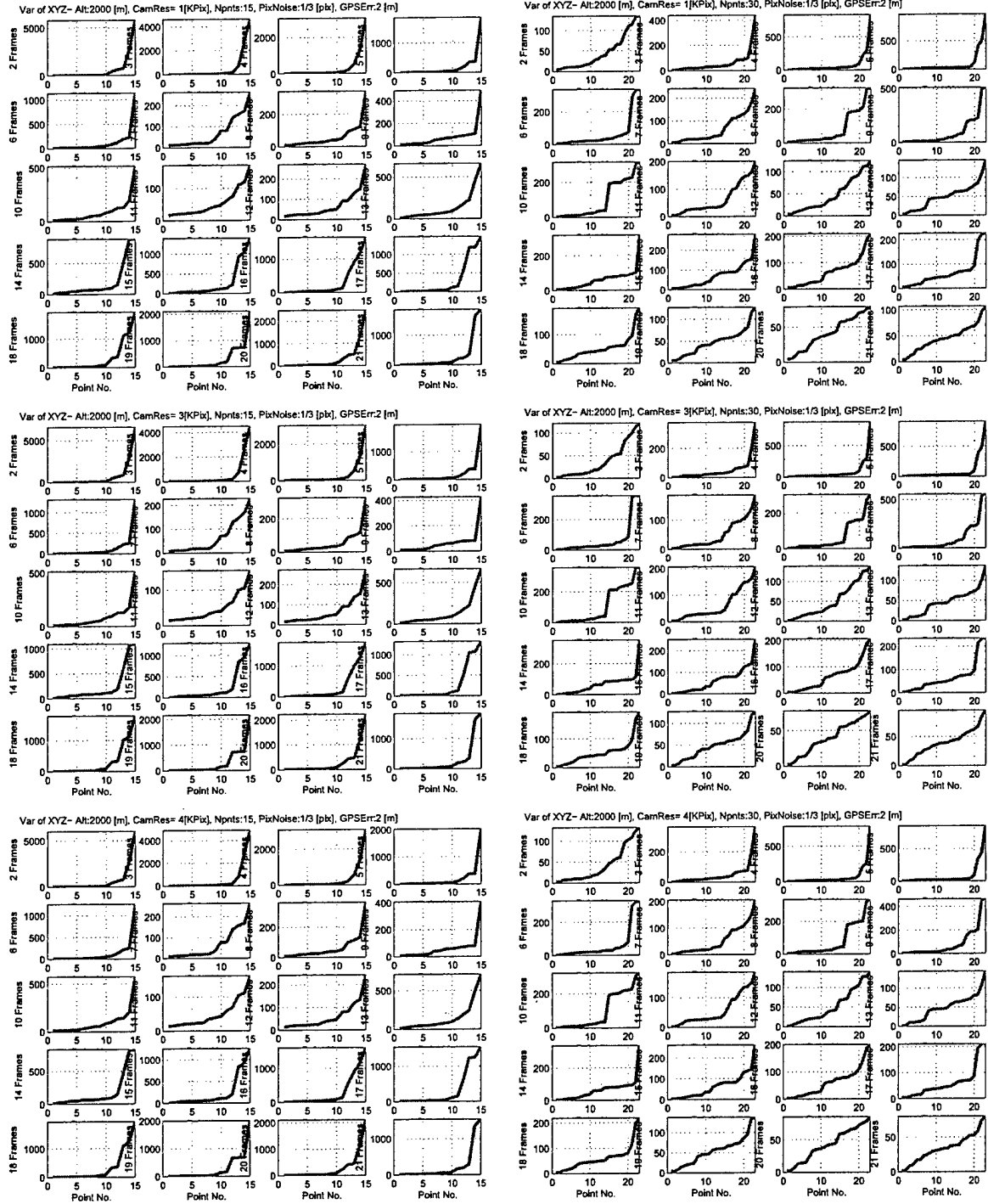


Figure 14: (continued)– Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

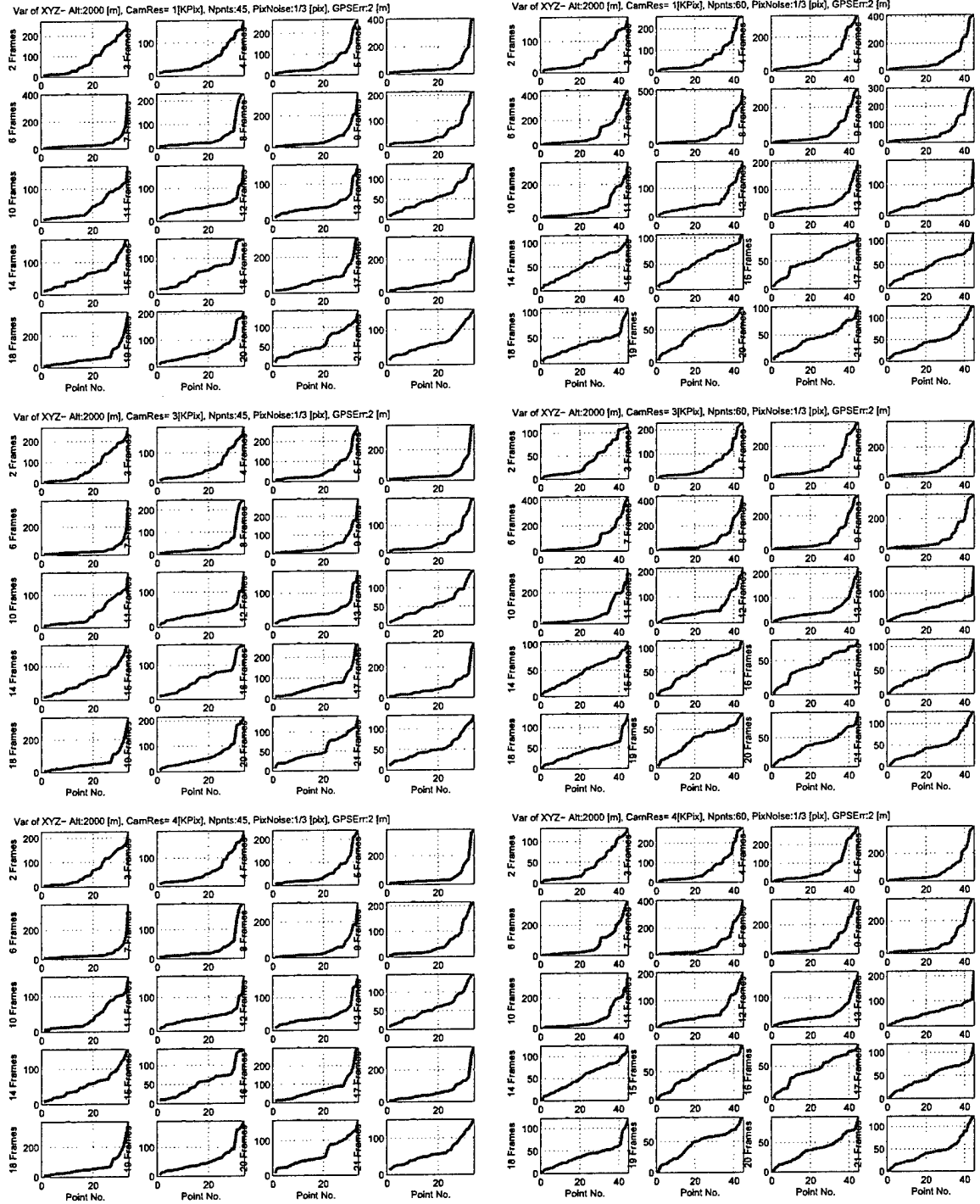


Figure 14: (continued)– Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

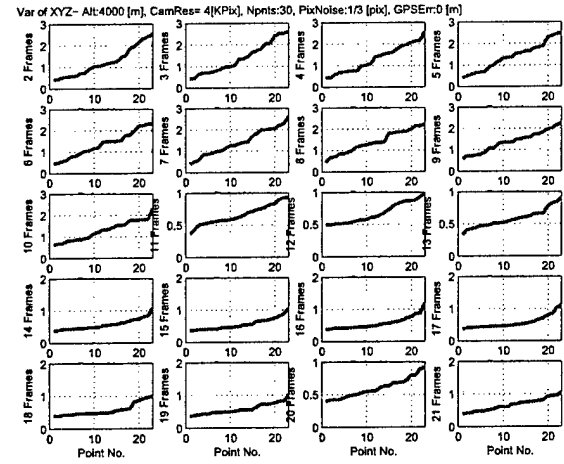
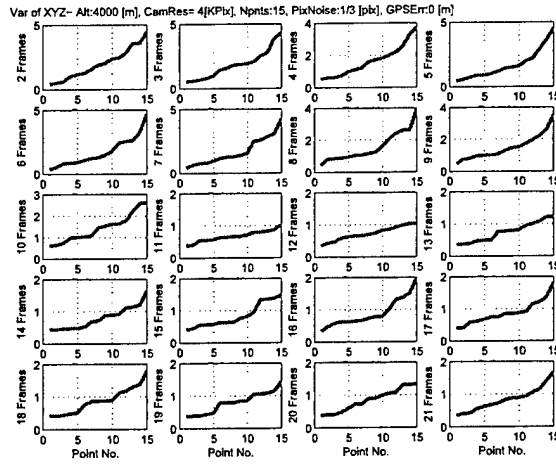
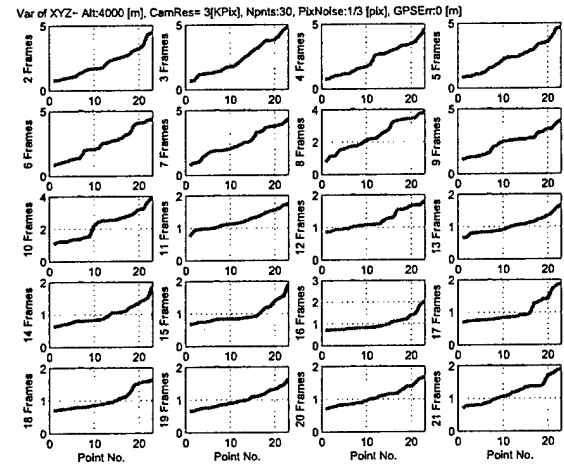
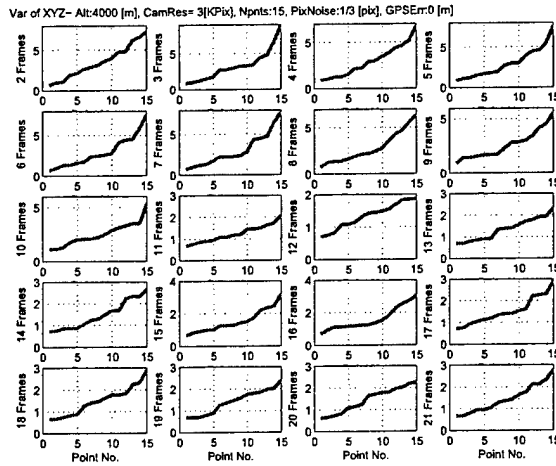
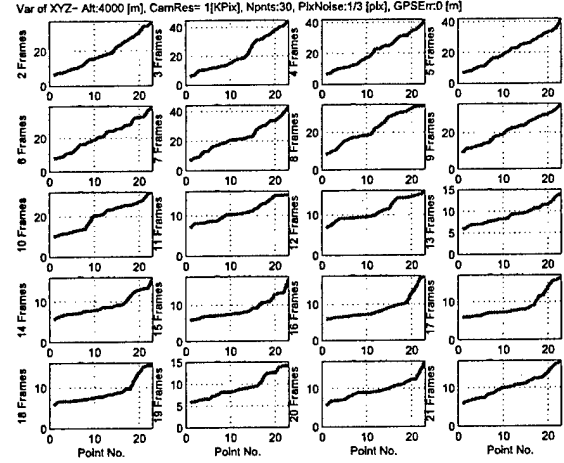
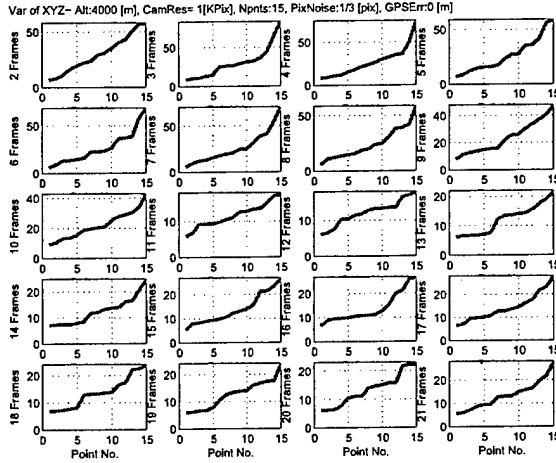


Figure 15: Closed-Form 8-Point Algorithm- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (altitude of 4000 [m]).

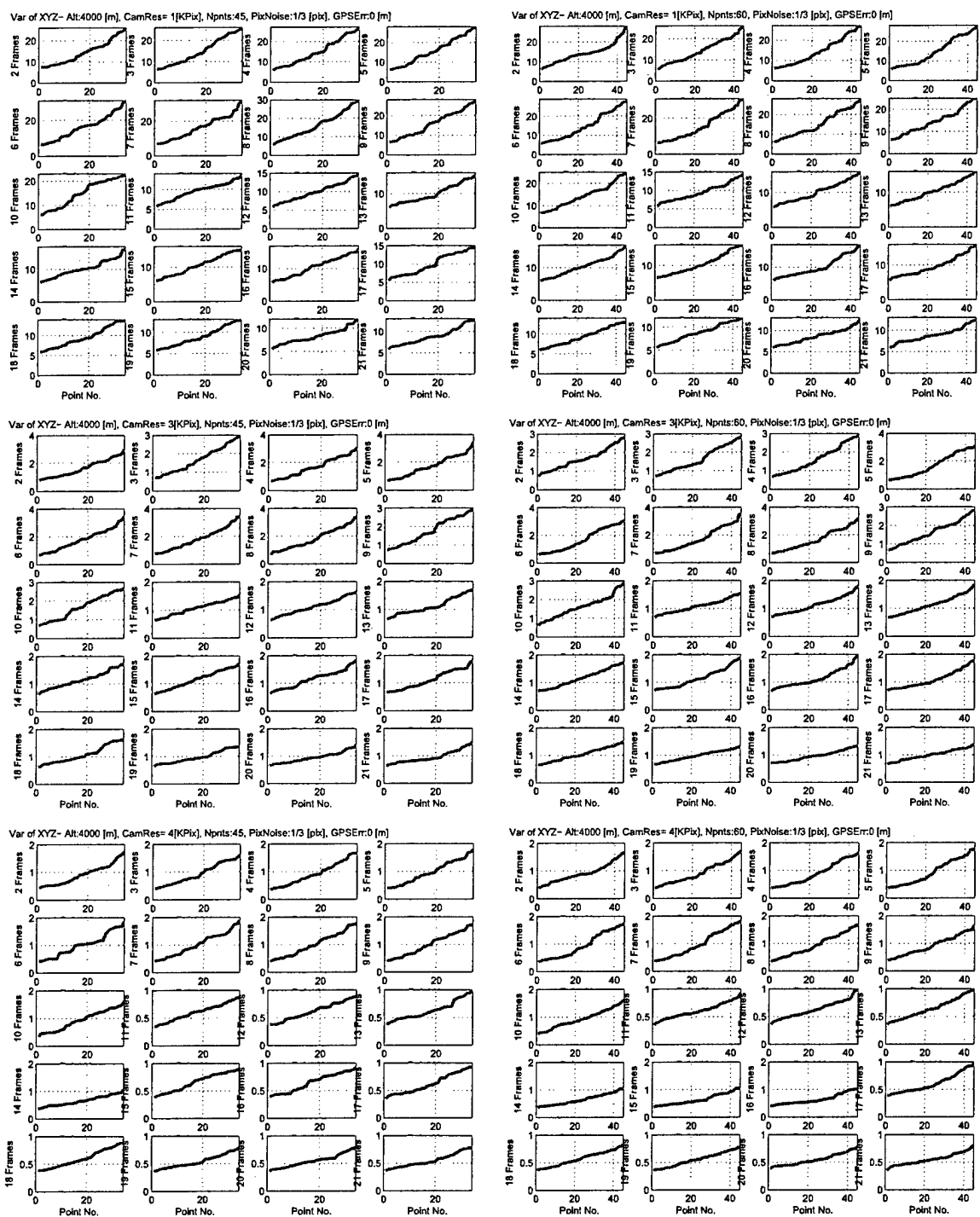


Figure 15: (continued)– Tracking 45 (left) and 60 (right) points with noise-free GPS.

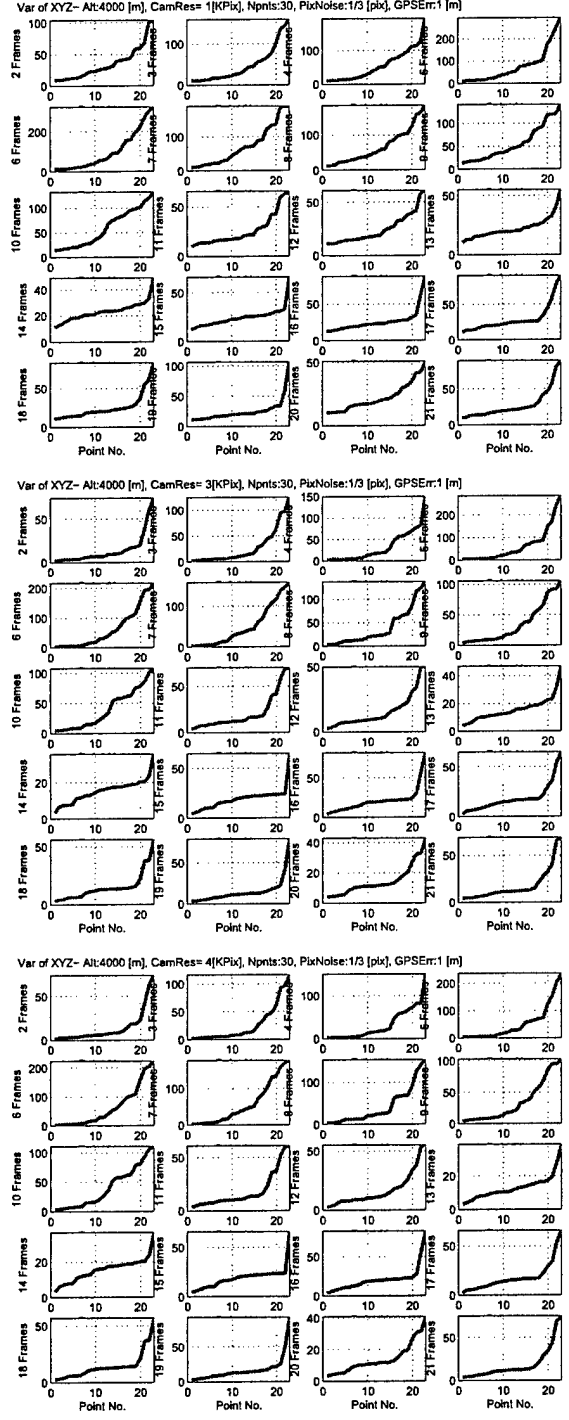
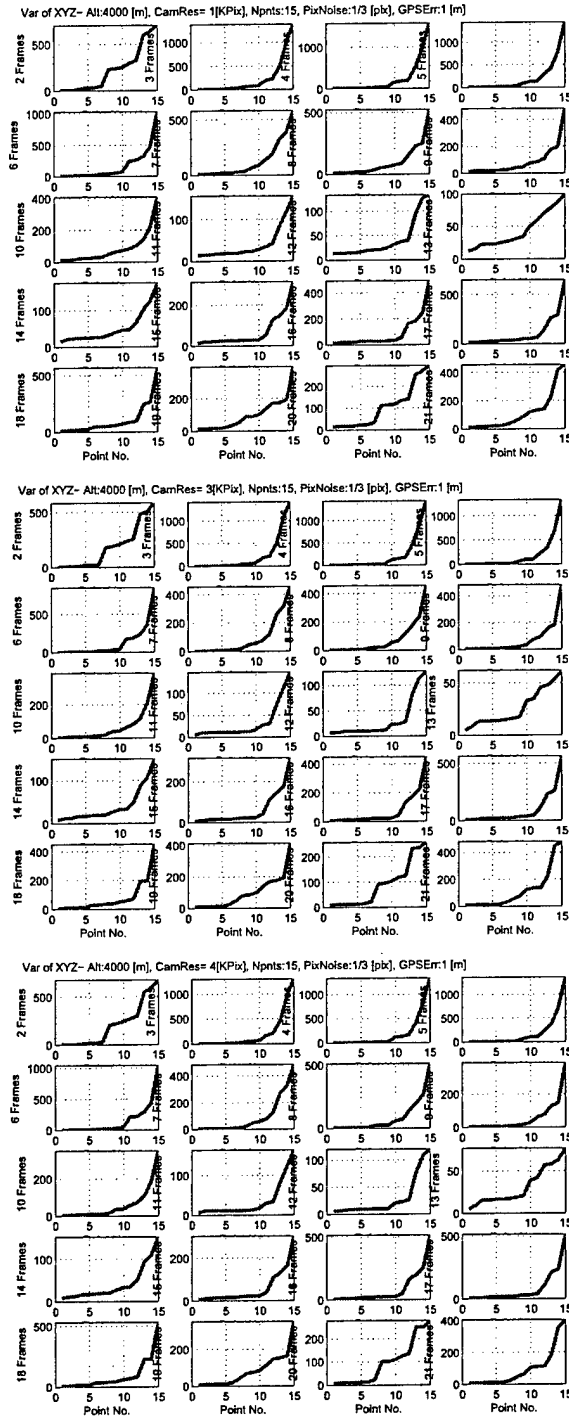


Figure 15: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

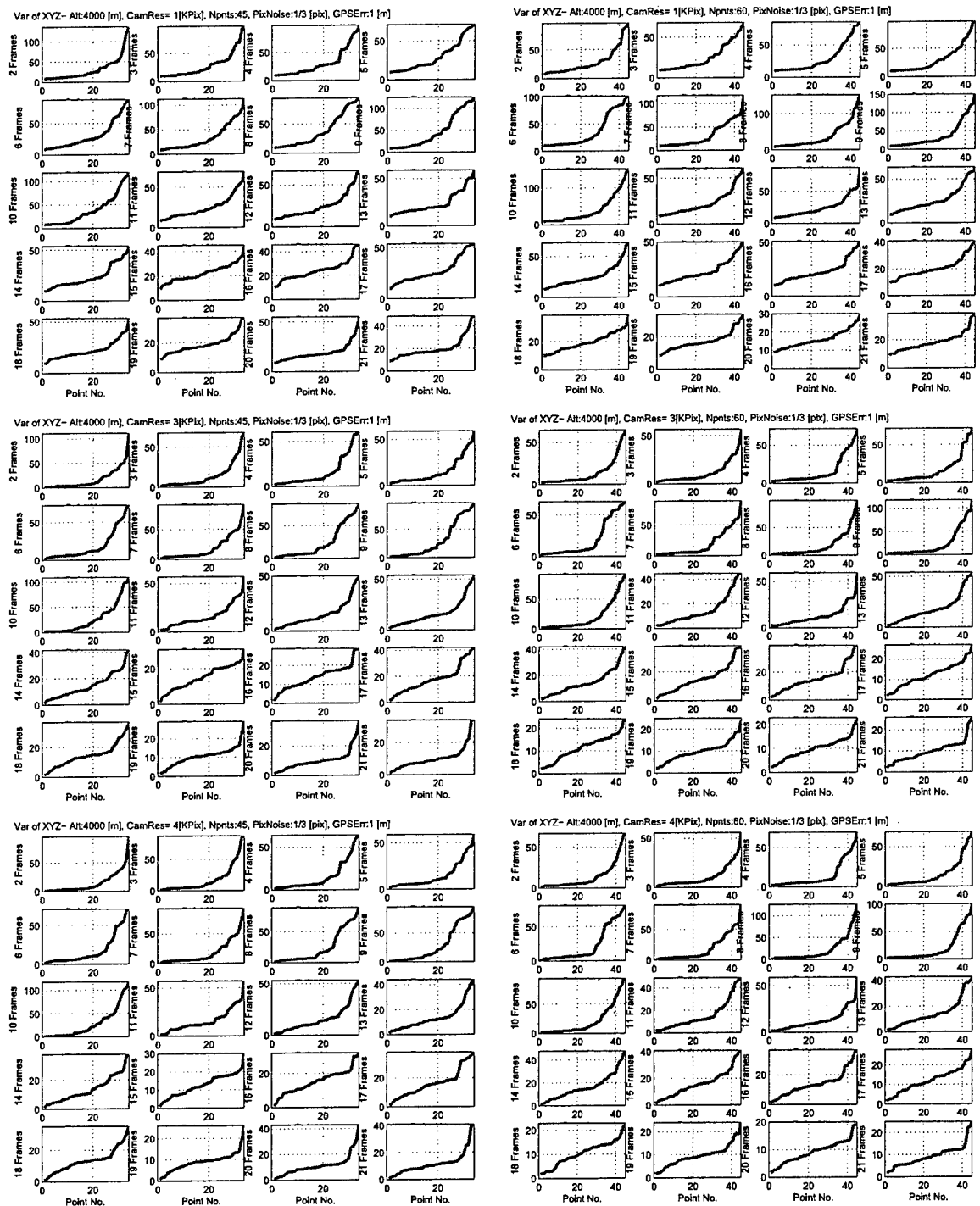


Figure 15: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

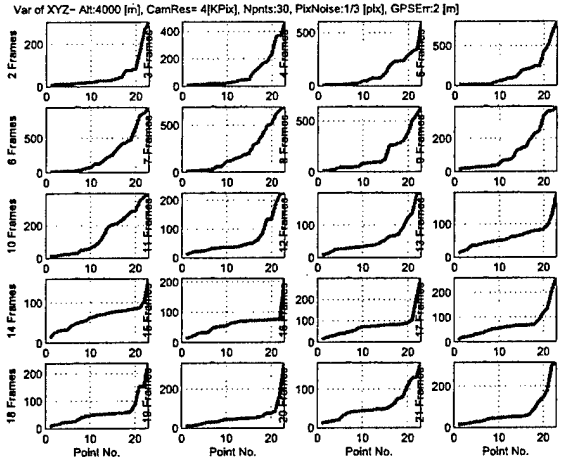
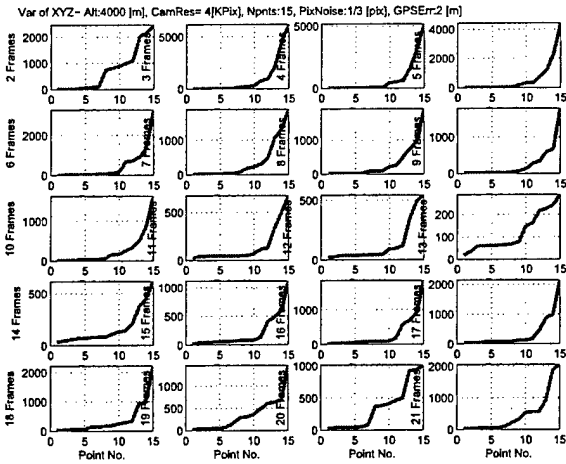
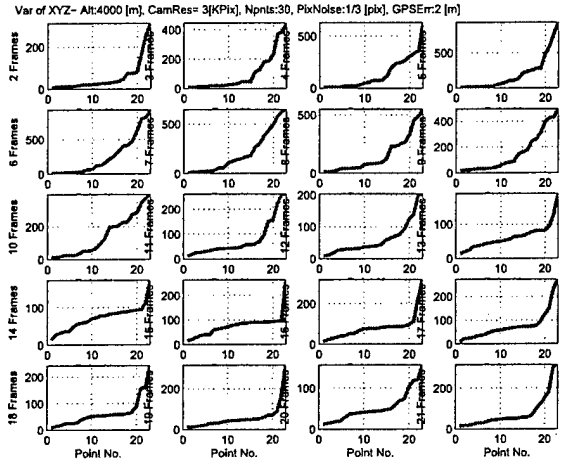
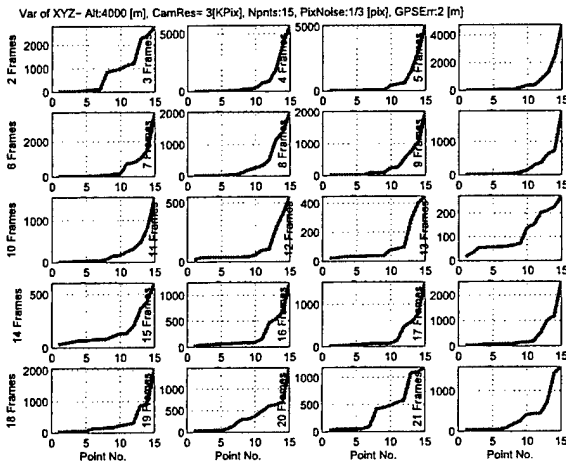
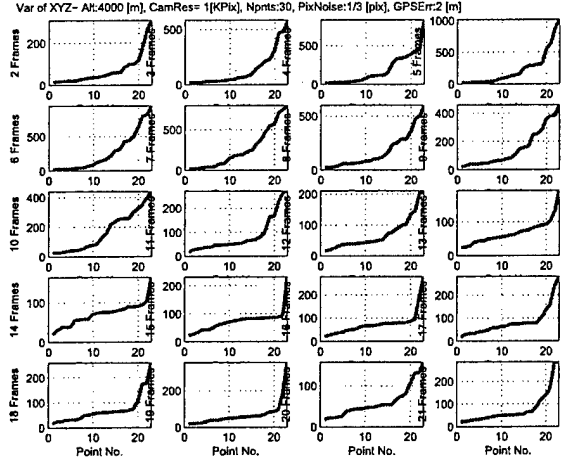
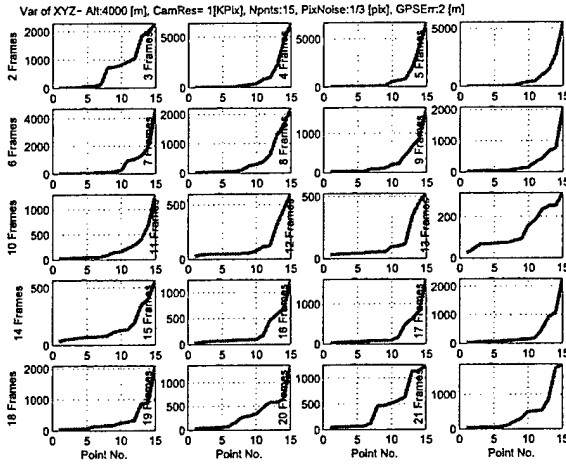


Figure 15: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

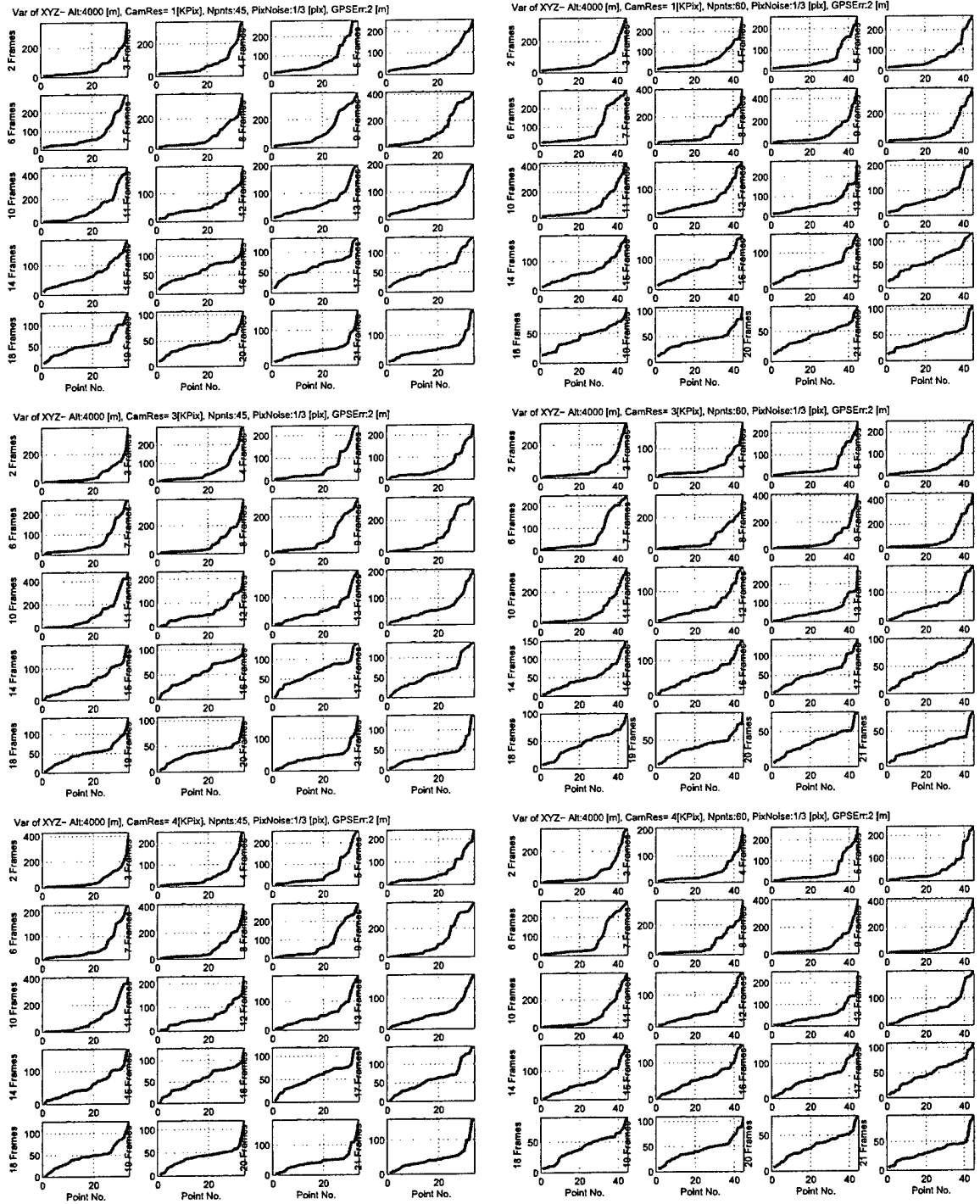


Figure 15: (continued)– Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

Appendix 3c: Results of closed-form solution with small-angle rotations for terrain feature localization

Results in this appendix are from simulations for the three altitudes of 500 [m], 2000 [m] and 4000 [m]. Each page consists of 5×4 arrays of plots arranged in 3 rows and 2 columns. Each row corresponds to one of the three $L \times L$ ($L = \{1, 3, 4\}$) image resolutions. Each of the two columns deal with one of N ($N = \{15, 30, 45, 60\}$) number of terrain features that are used in the computation of the UAV pose from two views. The complete set for all four choices for N are given on two subsequent pages. Each of the 5×4 arrays correspond to computations based on tracking the features in M ($M = 2 : 21$) views. Finally, various pages contain the results for GPS variances $\sigma_{GPS} = \{0, 1, 2\}$ [m].

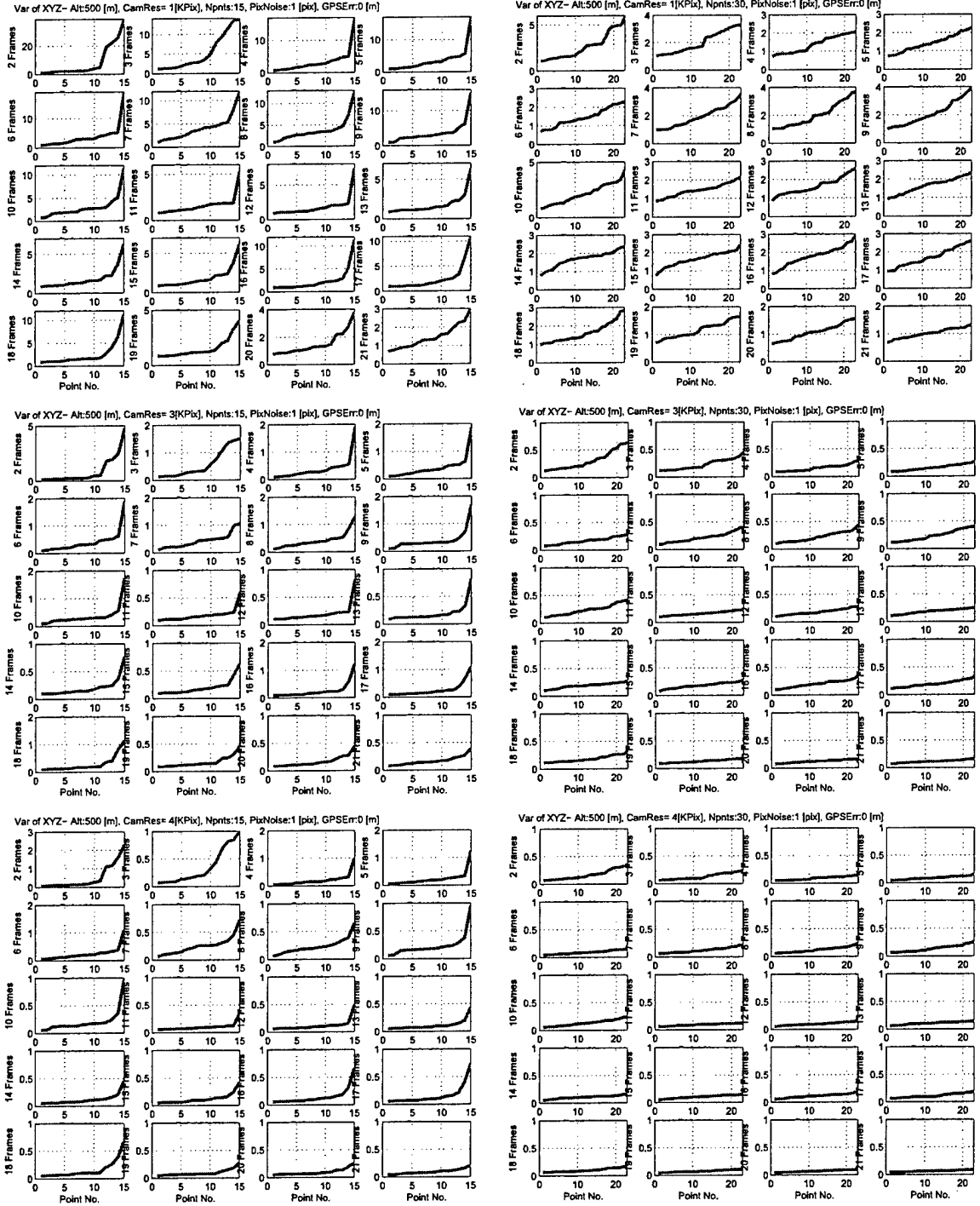


Figure 16: Closed-Form Solution with Small Rotation Approximation– Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 500); See text for details.

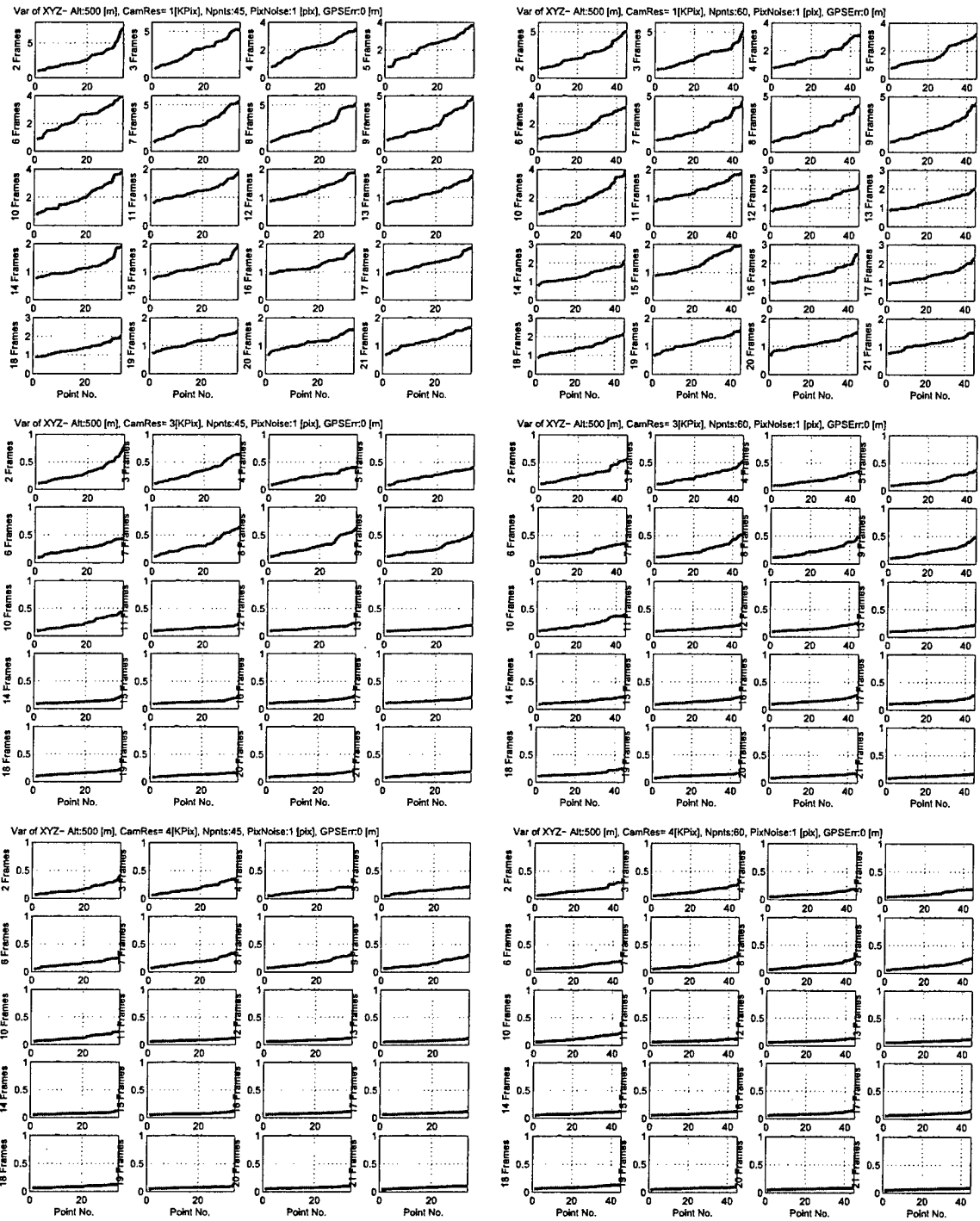


Figure 16: (continued)– Tracking 45 (left) and 60 (right) points with noise-free GPS.

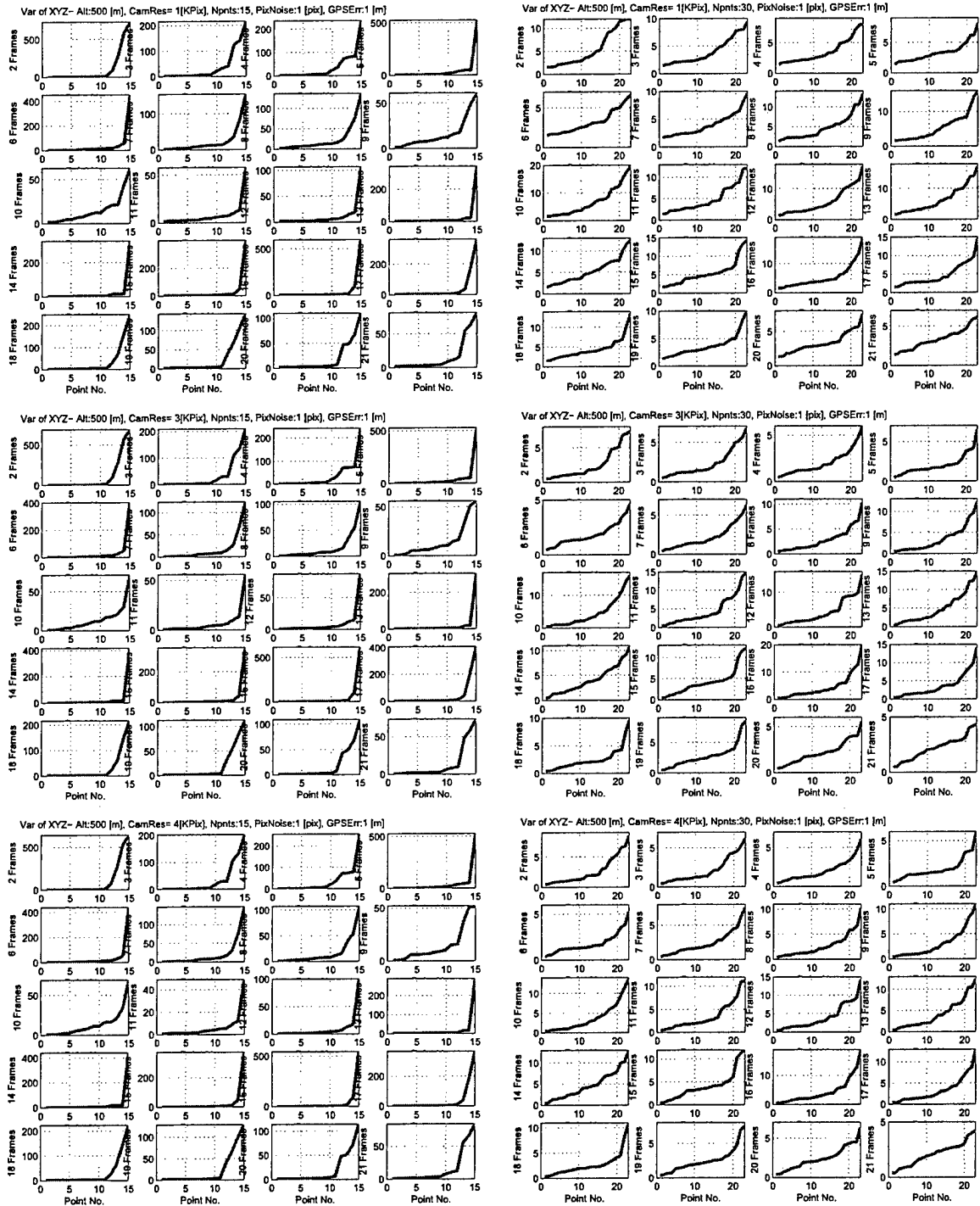


Figure 16: (continued)– Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

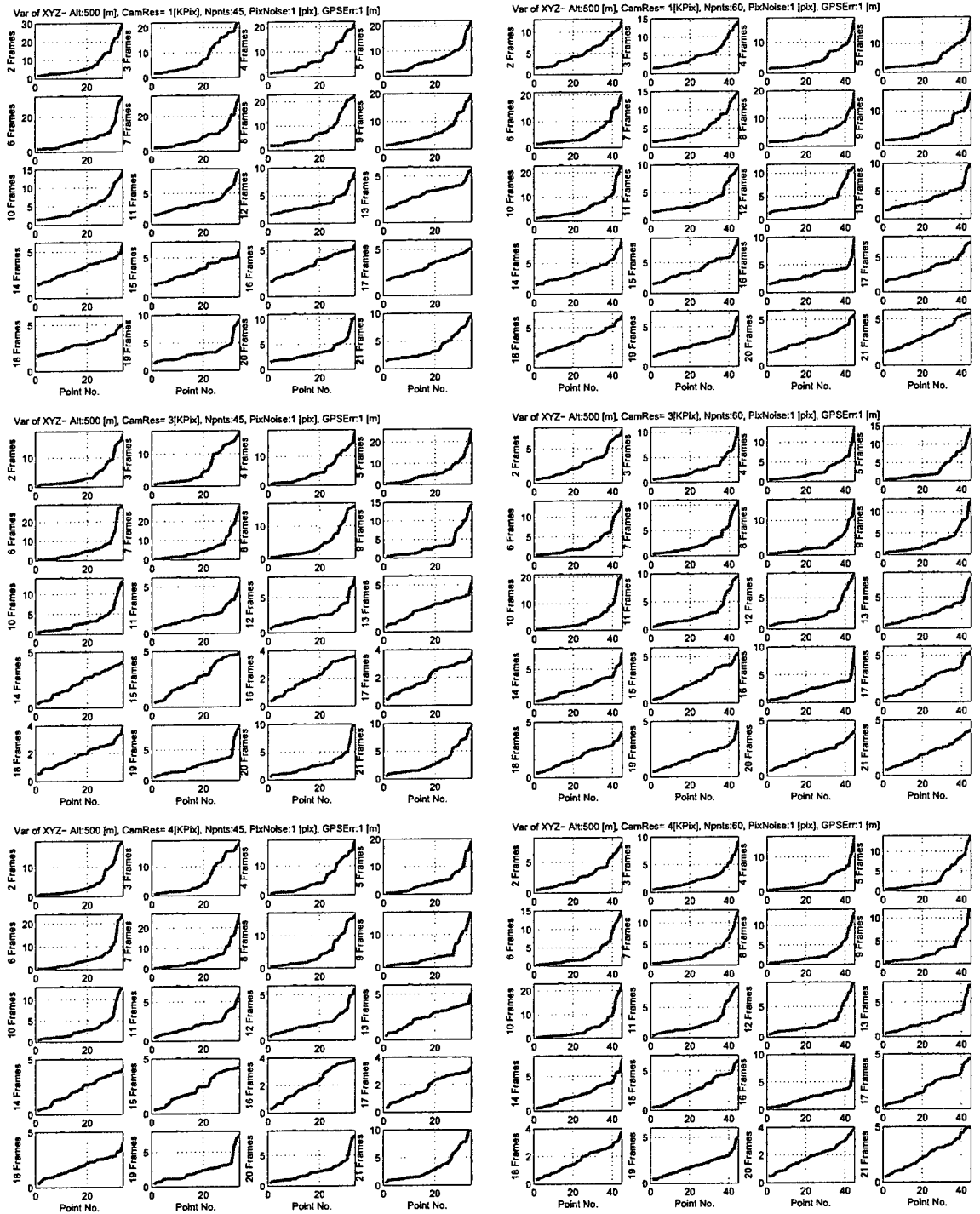


Figure 16: (continued)– Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

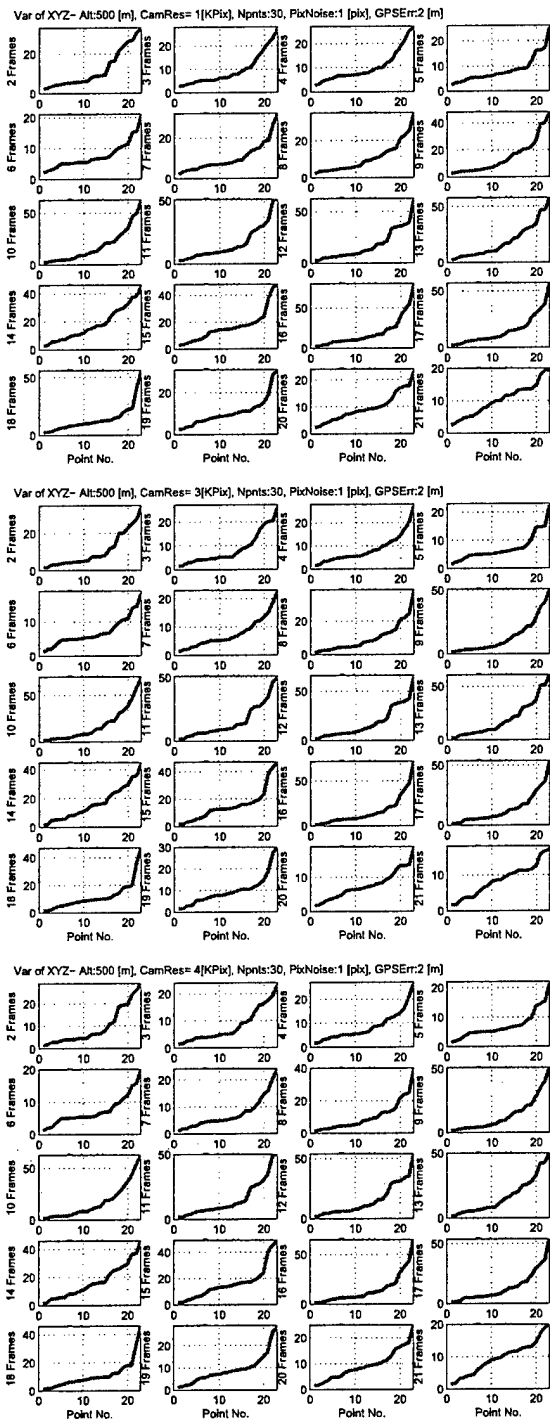
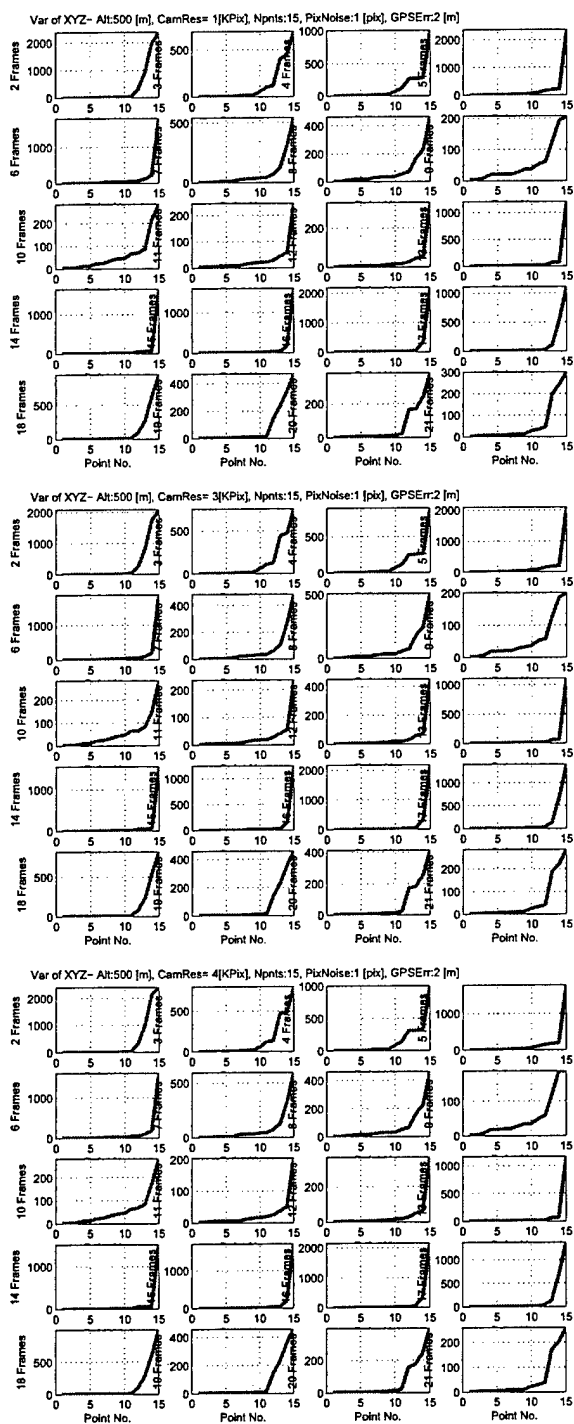


Figure 16: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

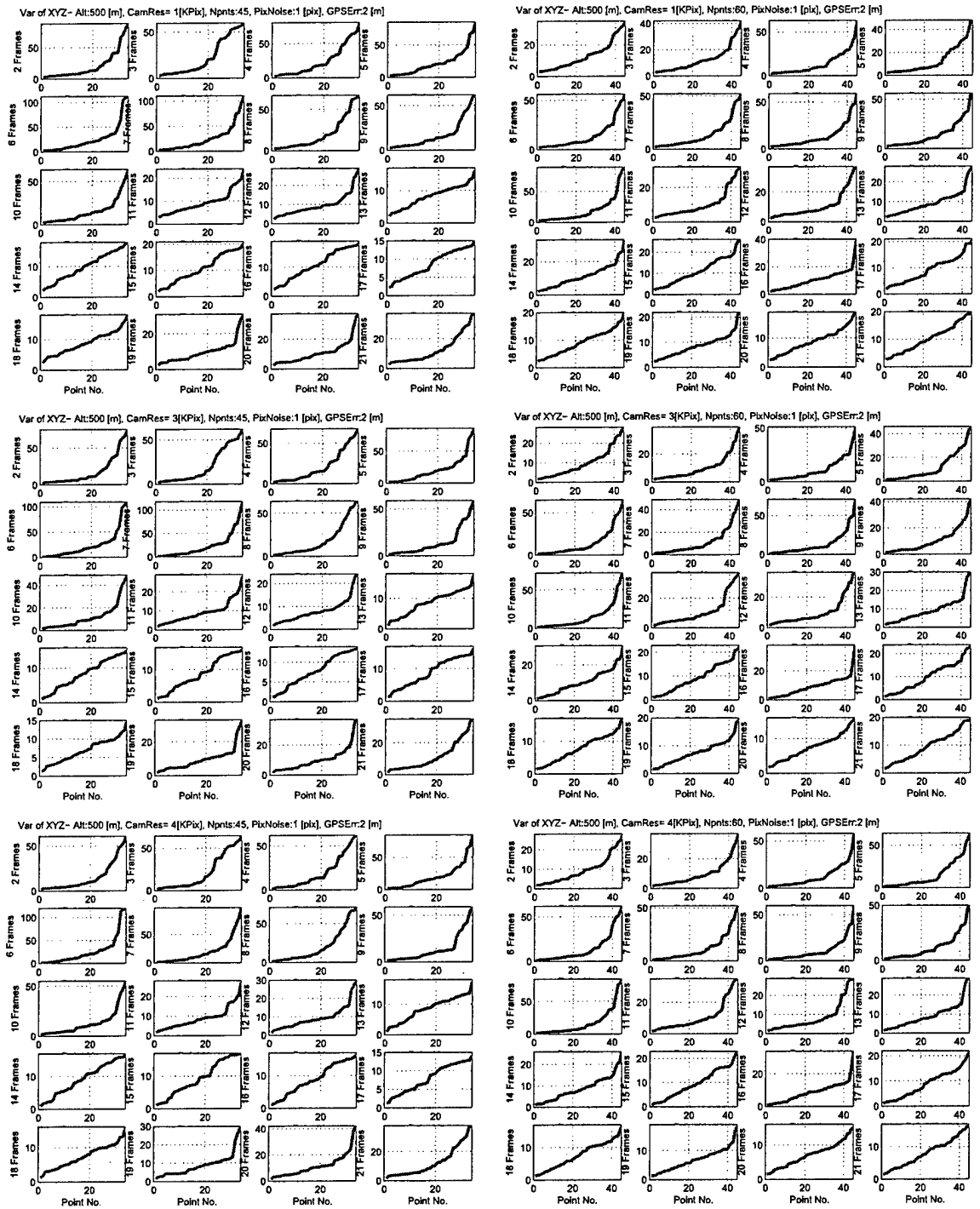
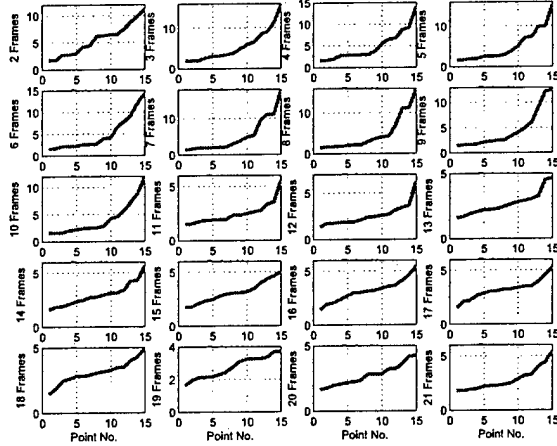
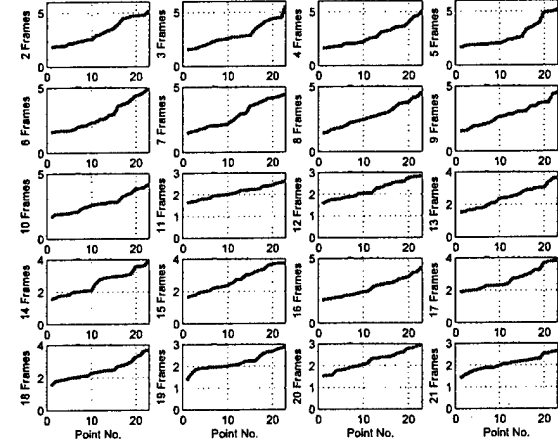


Figure 16: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

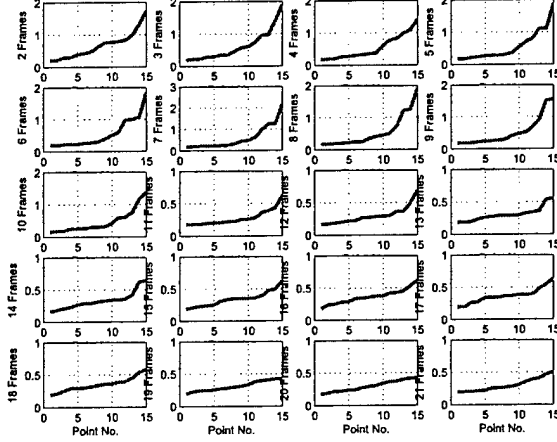
Var of XYZ- Alt:2000 [m], CamRes= 1[KPix], Npnts:15, PixNoise:1/3 [pix], GPSErr:0 [m]



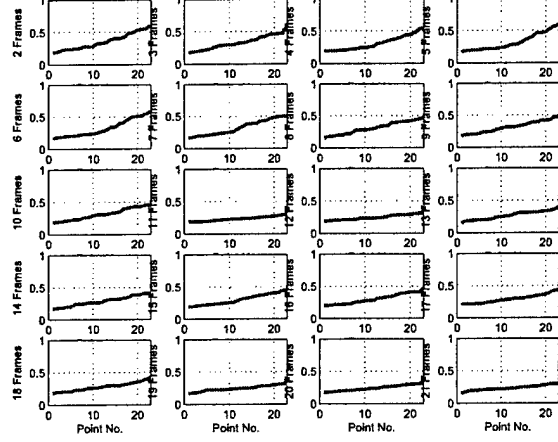
Var of XYZ- Alt:2000 [m], CamRes= 1[KPix], Npnts:30, PixNoise:1/3 [pix], GPSErr:0 [m]



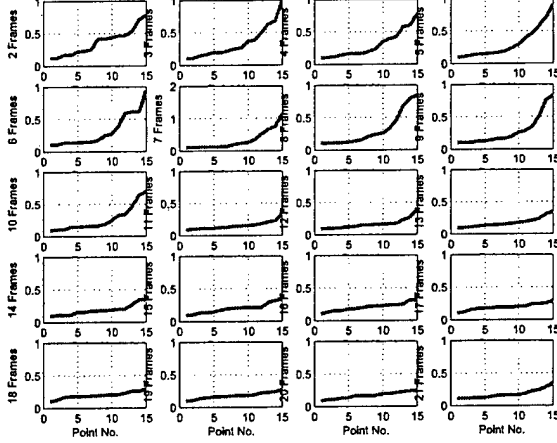
Var of XYZ- Alt:2000 [m], CamRes= 3[KPix], Npnts:15, PixNoise:1/3 [pix], GPSErr:0 [m]



Var of XYZ- Alt:2000 [m], CamRes= 3[KPix], Npnts:30, PixNoise:1/3 [pix], GPSErr:0 [m]



Var of XYZ- Alt:2000 [m], CamRes= 4[KPix], Npnts:15, PixNoise:1/3 [pix], GPSErr:0 [m]



Var of XYZ- Alt:2000 [m], CamRes= 4[KPix], Npnts:30, PixNoise:1/3 [pix], GPSErr:0 [m]

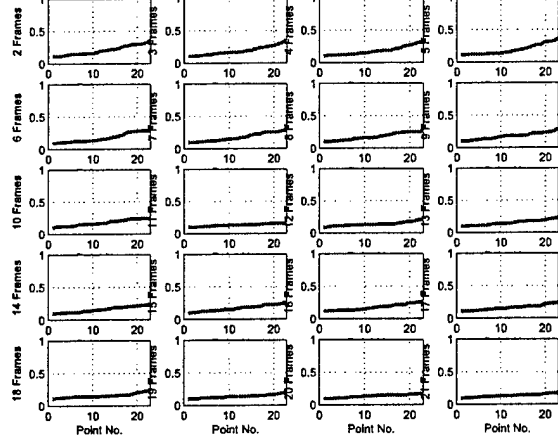


Figure 17: Closed-Form Solution with Small Rotation Approximation– Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 2000).

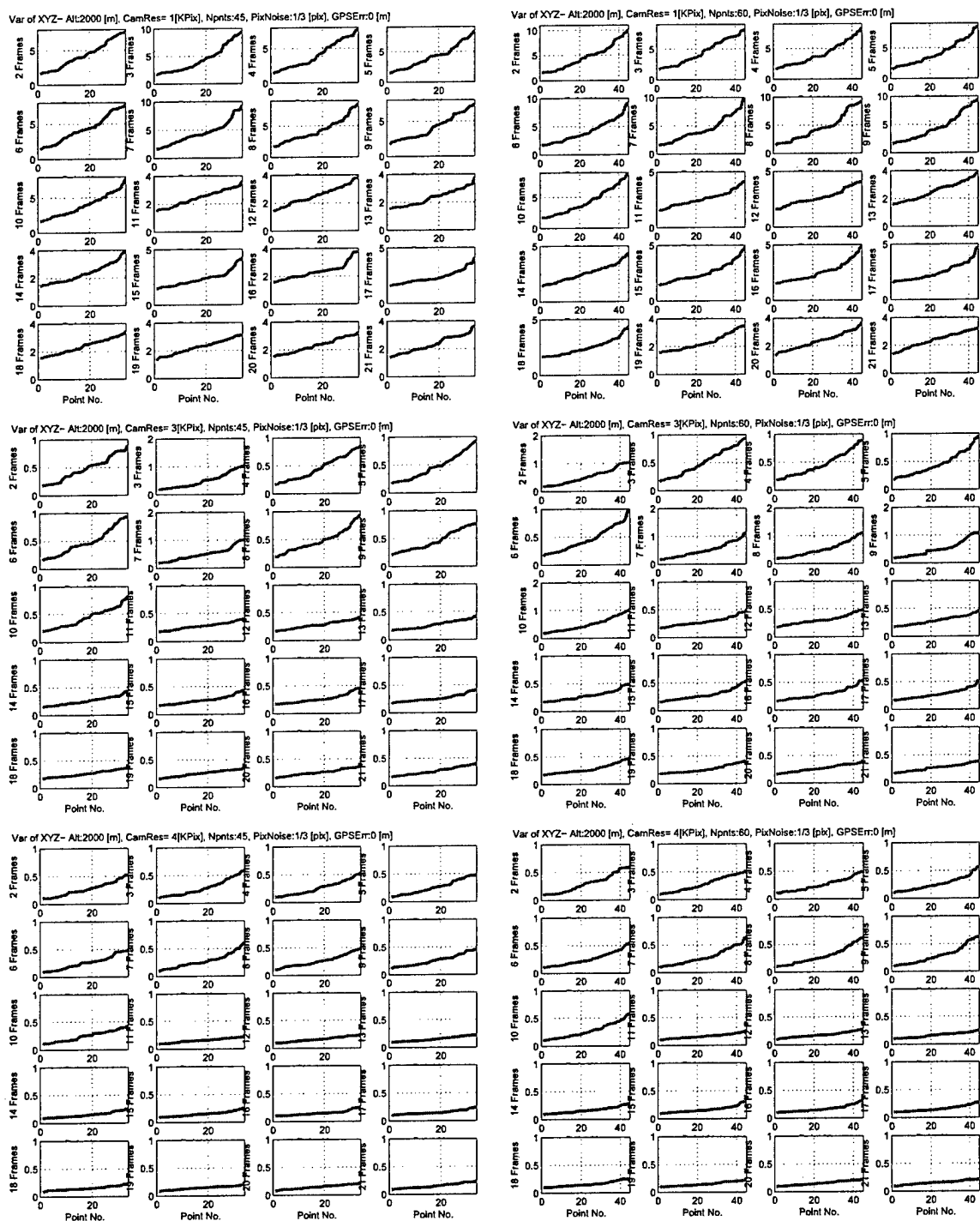


Figure 17: (continued)– Tracking 45 (left) and 60 (right) points with noise-free GPS.

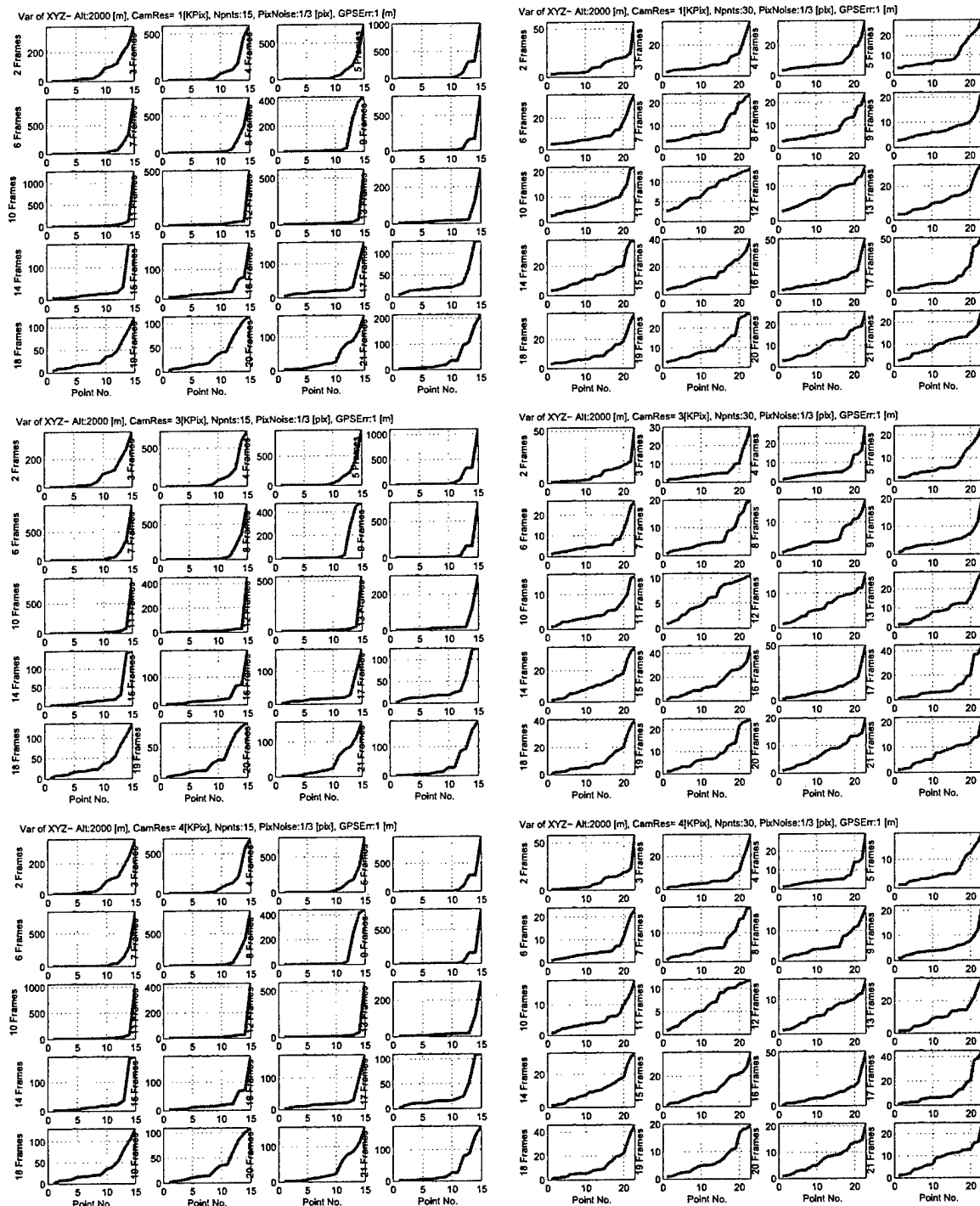


Figure 17: (continued)– Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

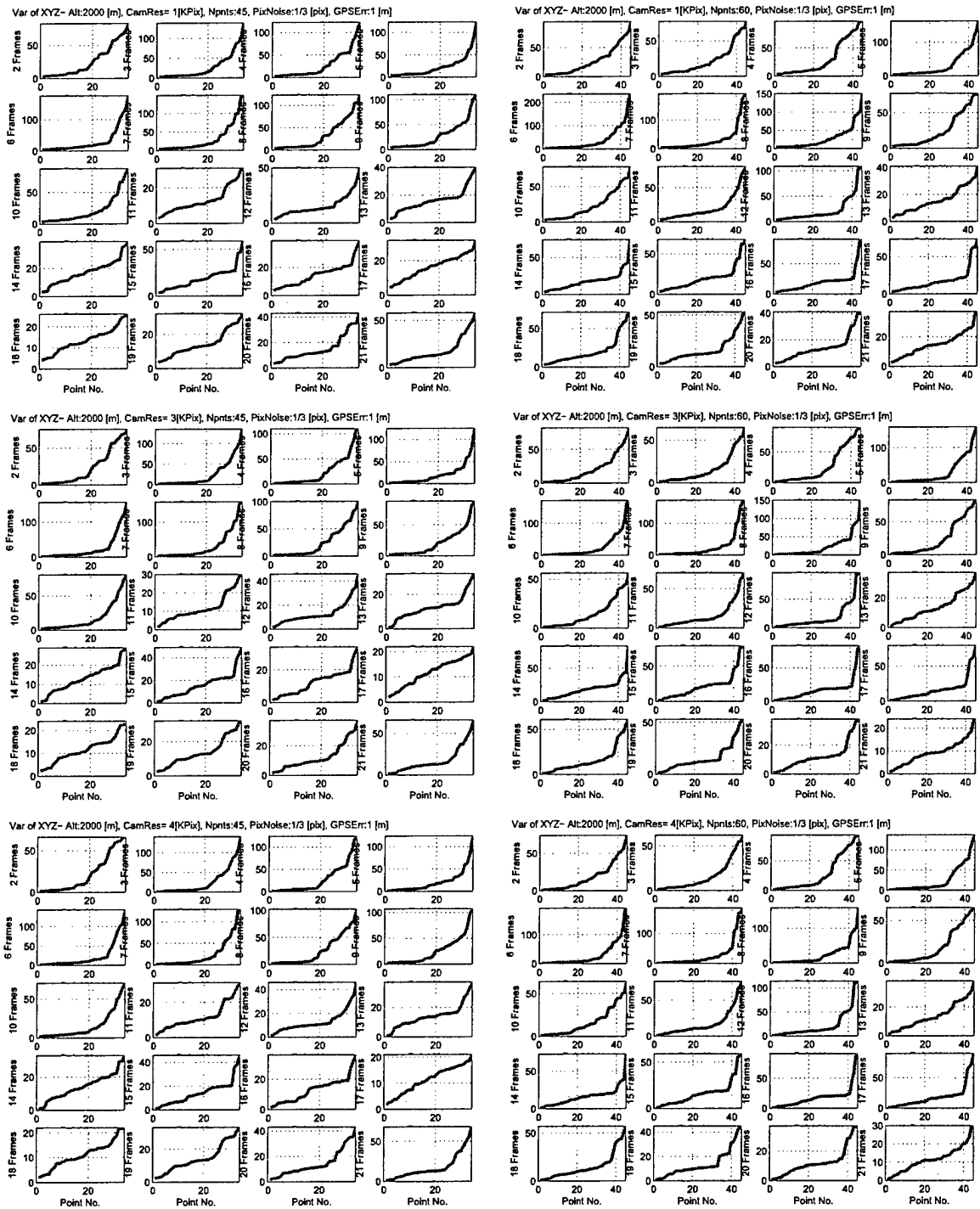


Figure 17: (continued)– Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

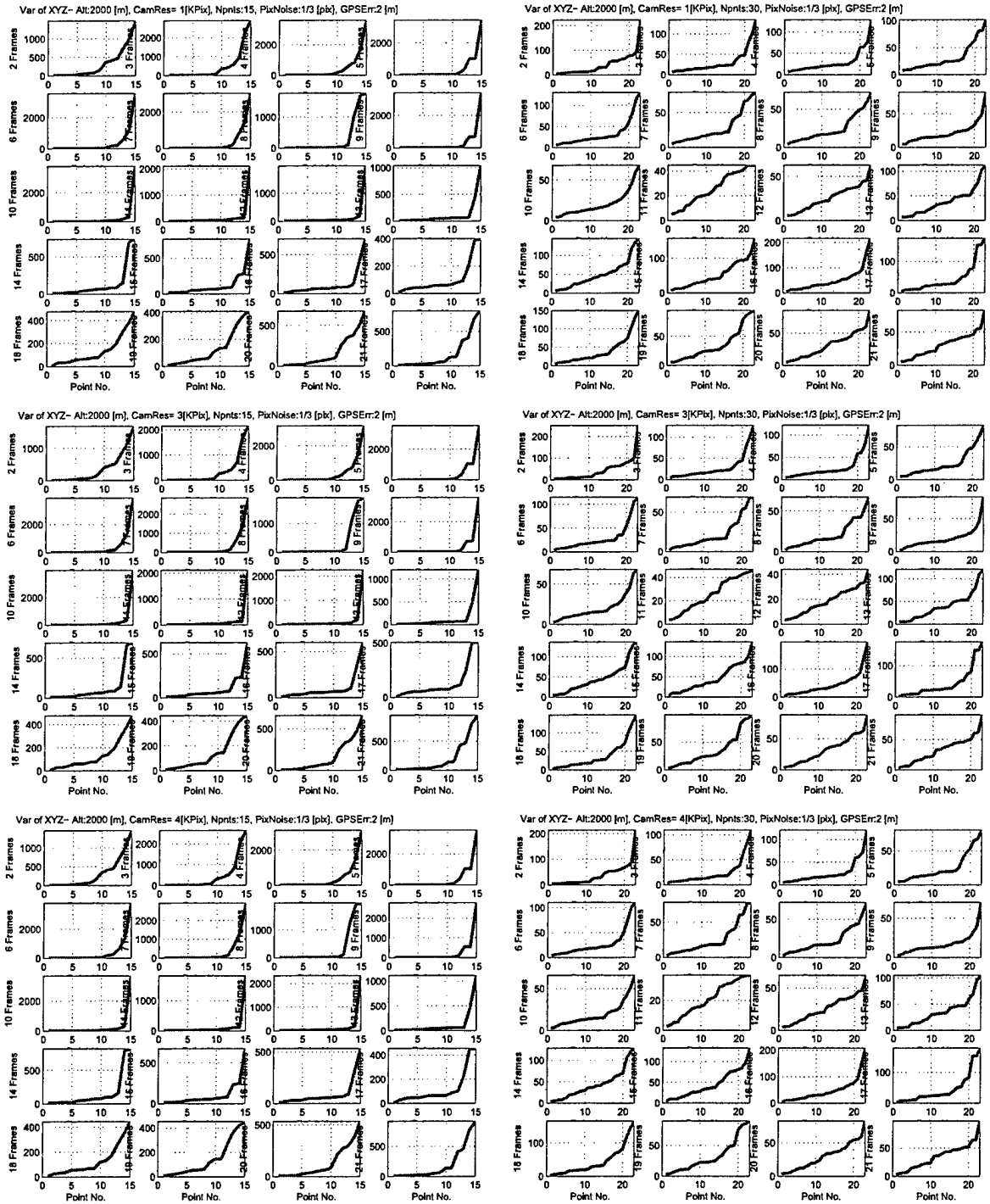


Figure 17: (continued)– Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

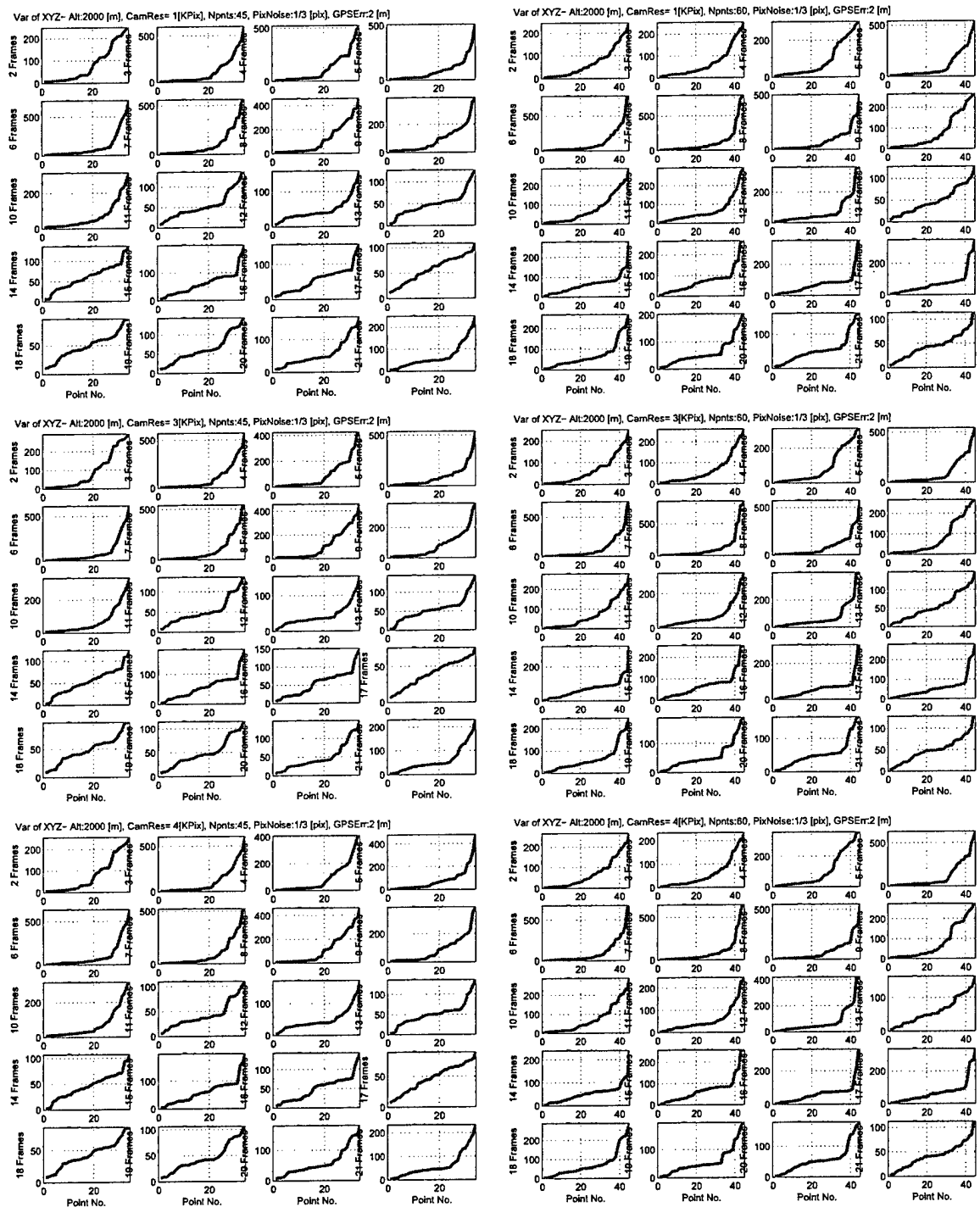


Figure 17: (continued)– Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

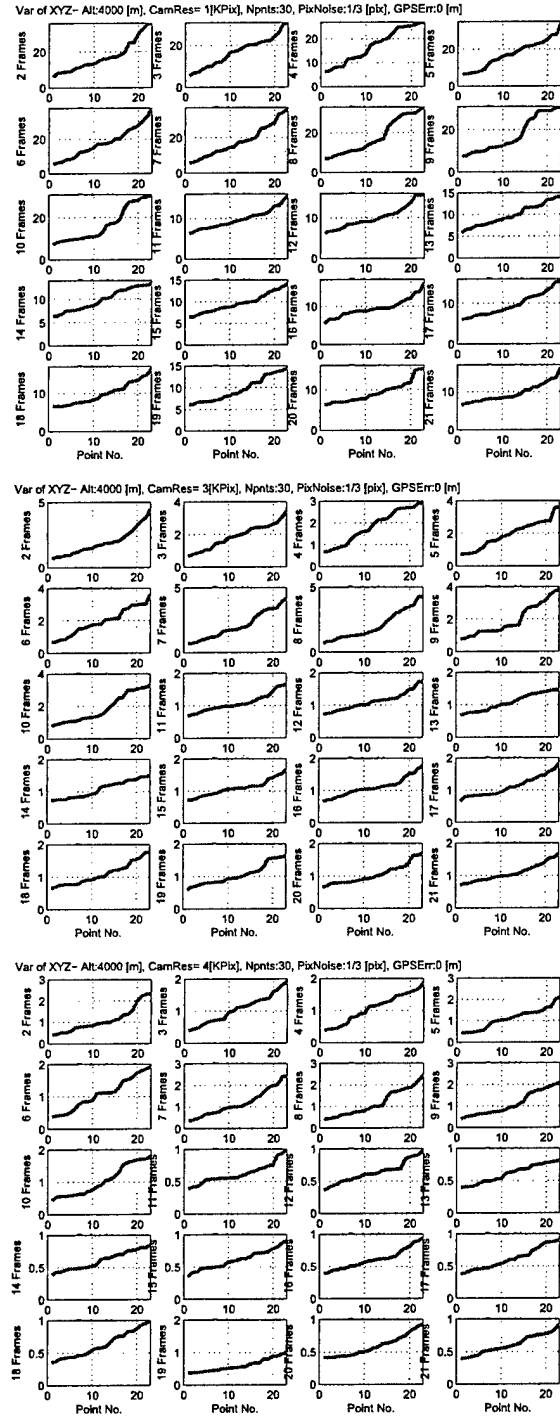
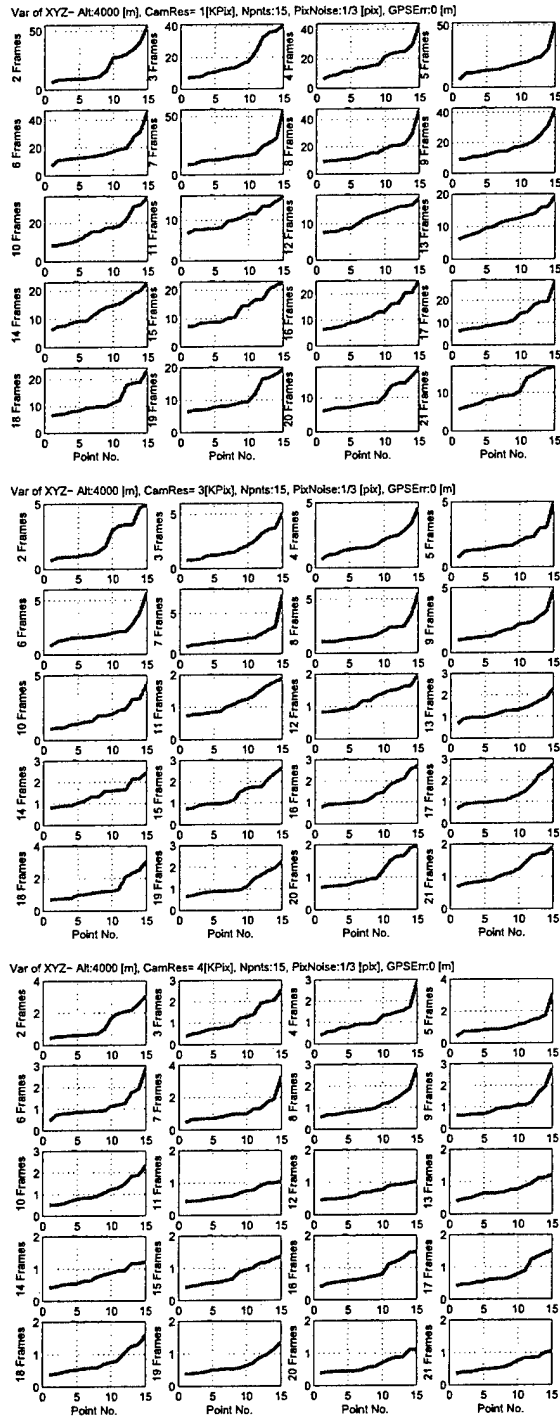


Figure 18: Closed-Form Solution with Small Rotation Approximation– Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 15 (left) and 30 (right) points with noise-free GPS (Altitude 4000).

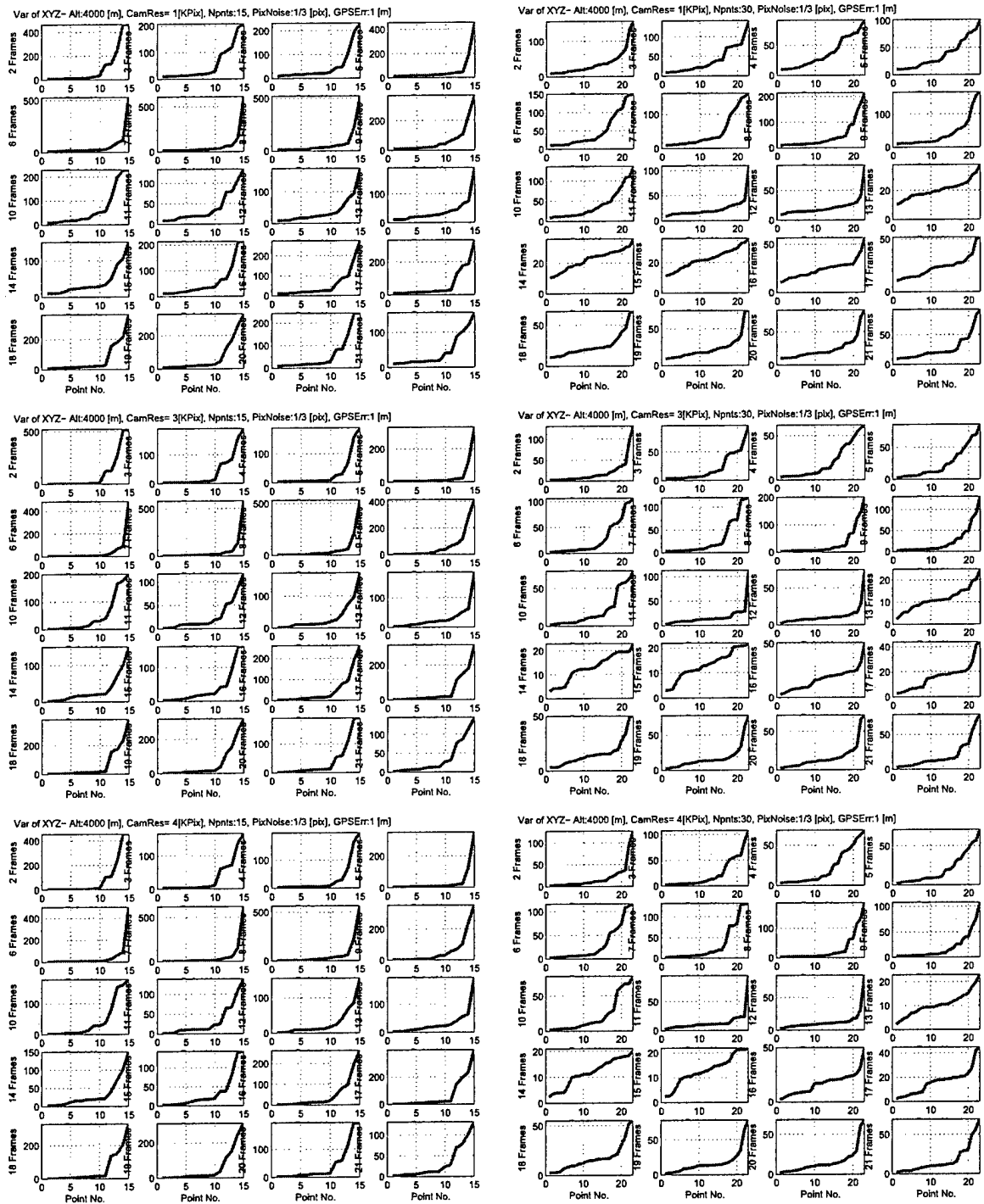


Figure 18: (continued)- Tracking 15 (left) and 30 (right) points with GPS error variance of 1 [m].

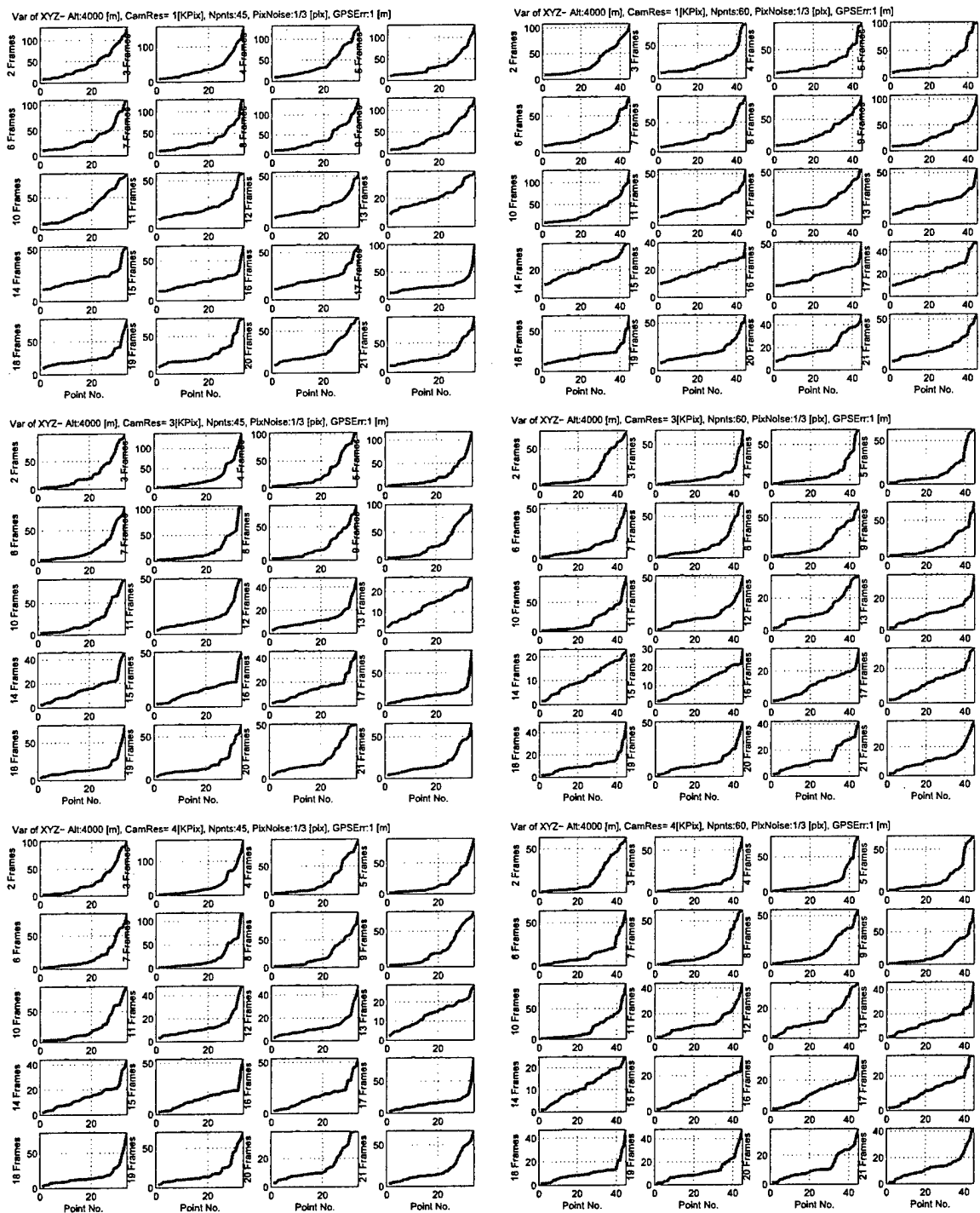


Figure 18: (continued)- Tracking 45 (left) and 60 (right) points with GPS error variance of 1 [m].

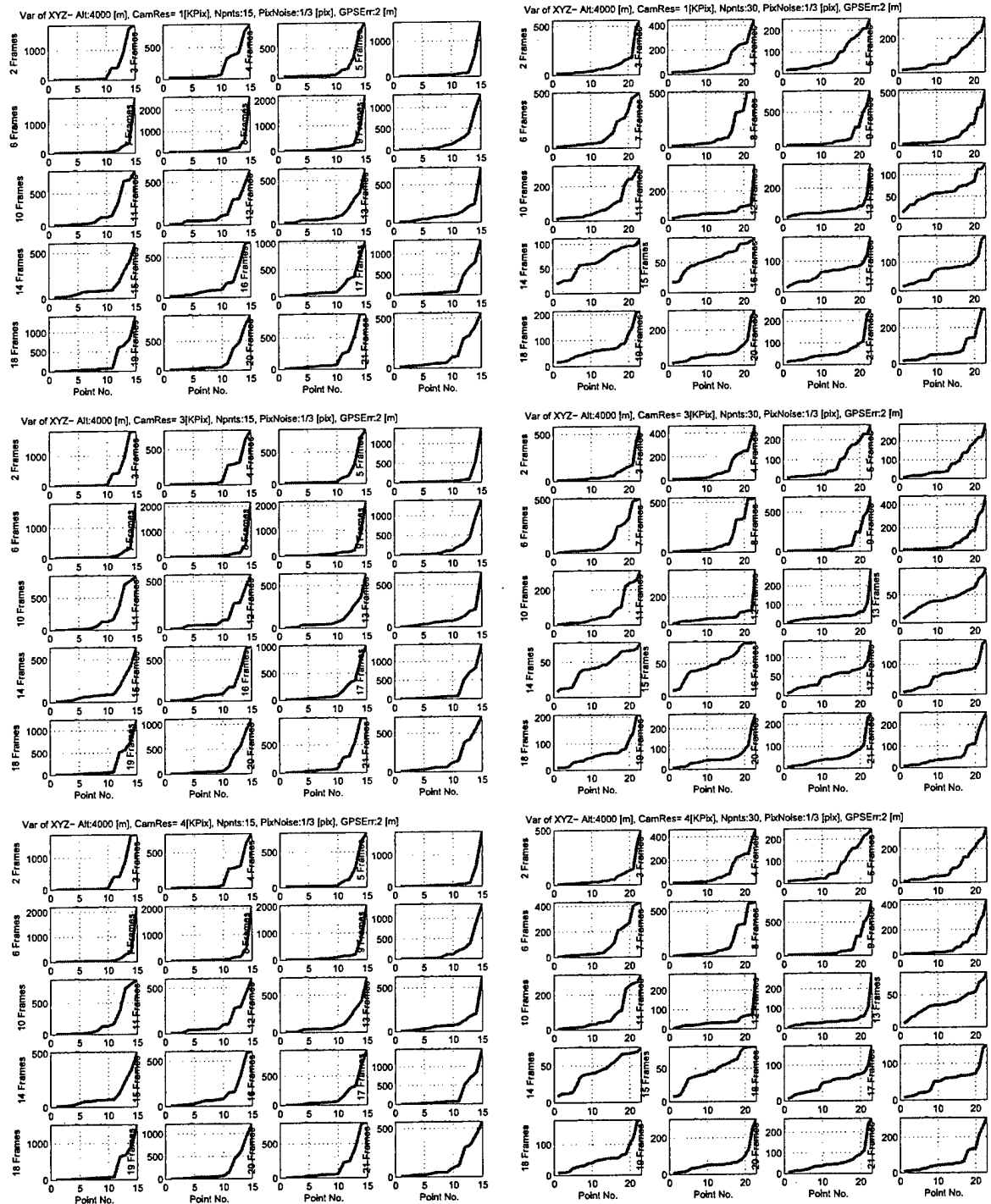


Figure 18: (continued)– Tracking 15 (left) and 30 (right) points with GPS error variance of 2 [m].

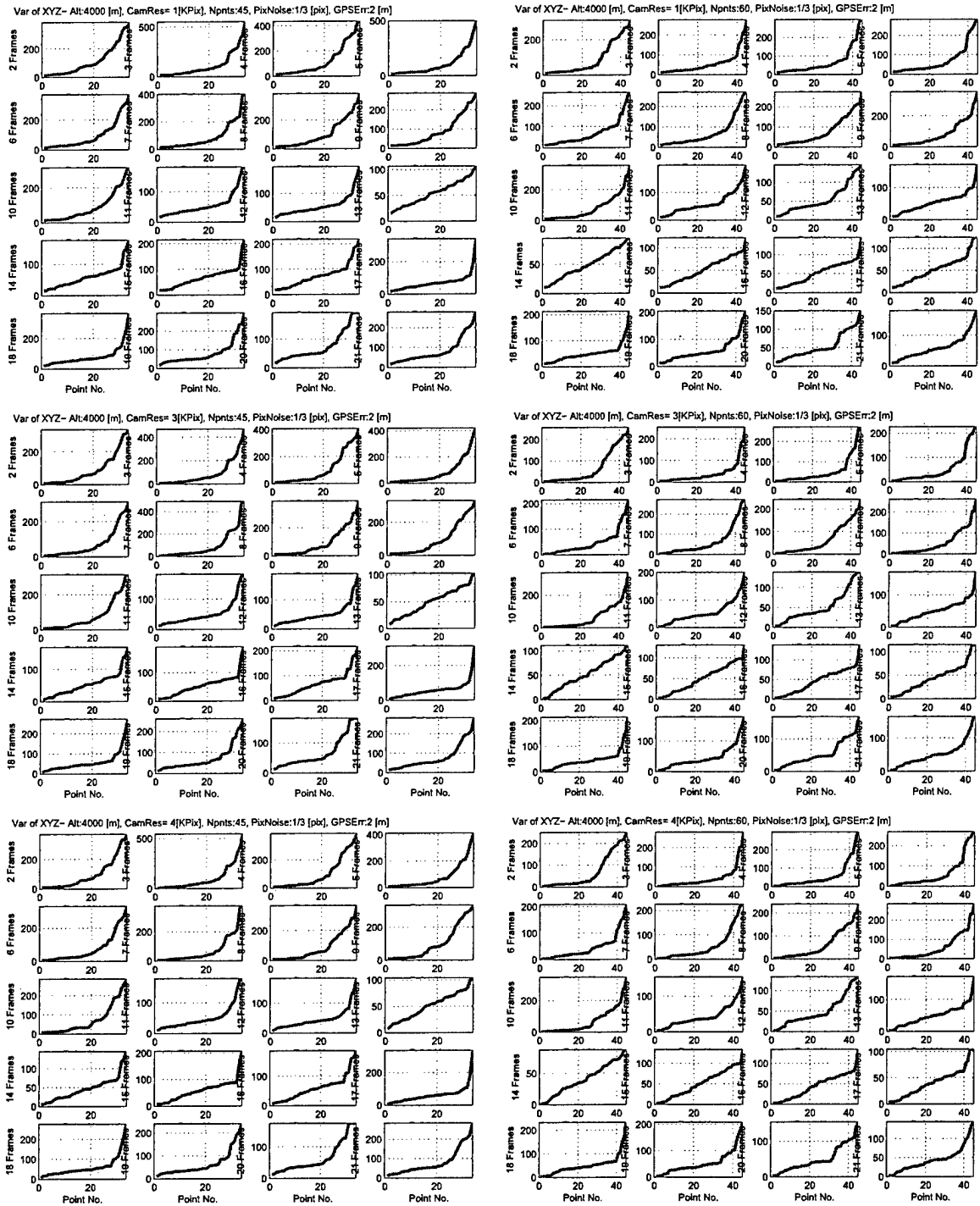


Figure 18: (continued)– Tracking 45 (left) and 60 (right) points with GPS error variance of 2 [m].

Appendix 3d: Selected results from nonlinear iterative solution for terrain feature localization

Each page in this section consists of six- arranged in 3 rows and 2 columns- 3×3 arrays of plots. Each row with 2 such arrays corresponds to one of three $L \times L$ ($L = \{1, 3, 4\}$) image resolutions. Each column of 3 such arrays comprises the results for one case of N ($N = \{3, 6, 9, 12\}$) terrain features being tracked (in $M = \{2 : 15\}$ frames) to determine the 3-D coordinates of landmark terrain features. The complete set for all four values of N are given on two subsequent pages. Each 3×3 array corresponds to error variances of X , Y and Z coordinates for a *sample of 3 out of N terrain features* (complete set is given in Appendix 3e). Each plot has three curves (R,G, and B colors) for GPS variances $\sigma_{GPS} = \{0, 1, 2\}$ [m], respectively. Finally, various pages contain the results for 3 altitudes of 500 [m], 2000 [m], and 4000 [m].

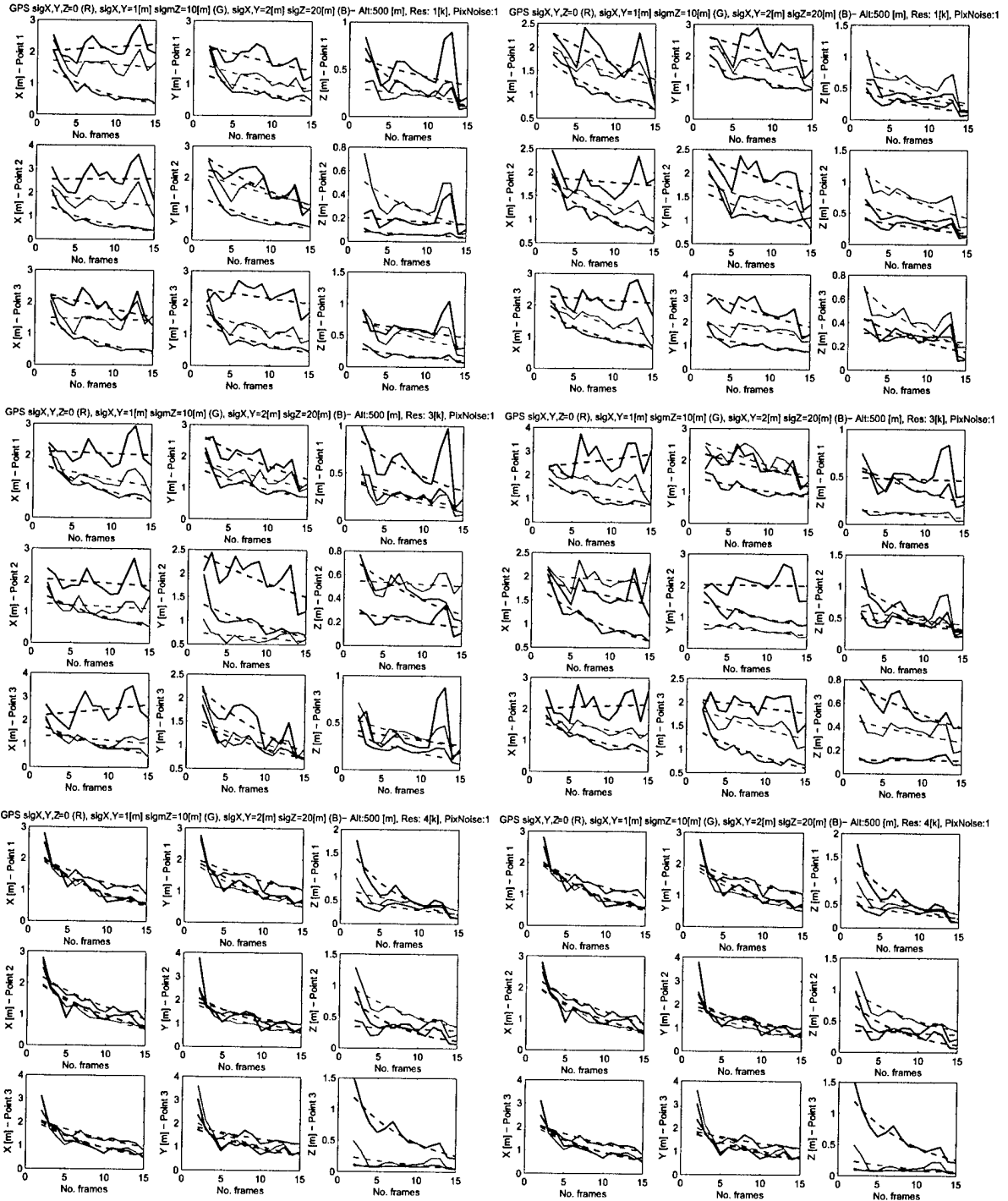


Figure 19: Nonlinear Iterative Solution- Uncertainty (reconstruction variance [m]) of 3 sample points in terrain feature localization by tracking 3 (left) and 6 (right) points in images with $L \times L$ ($L = \{1, 3, 4\}$) resolutions (Altitude 500); See text for details.

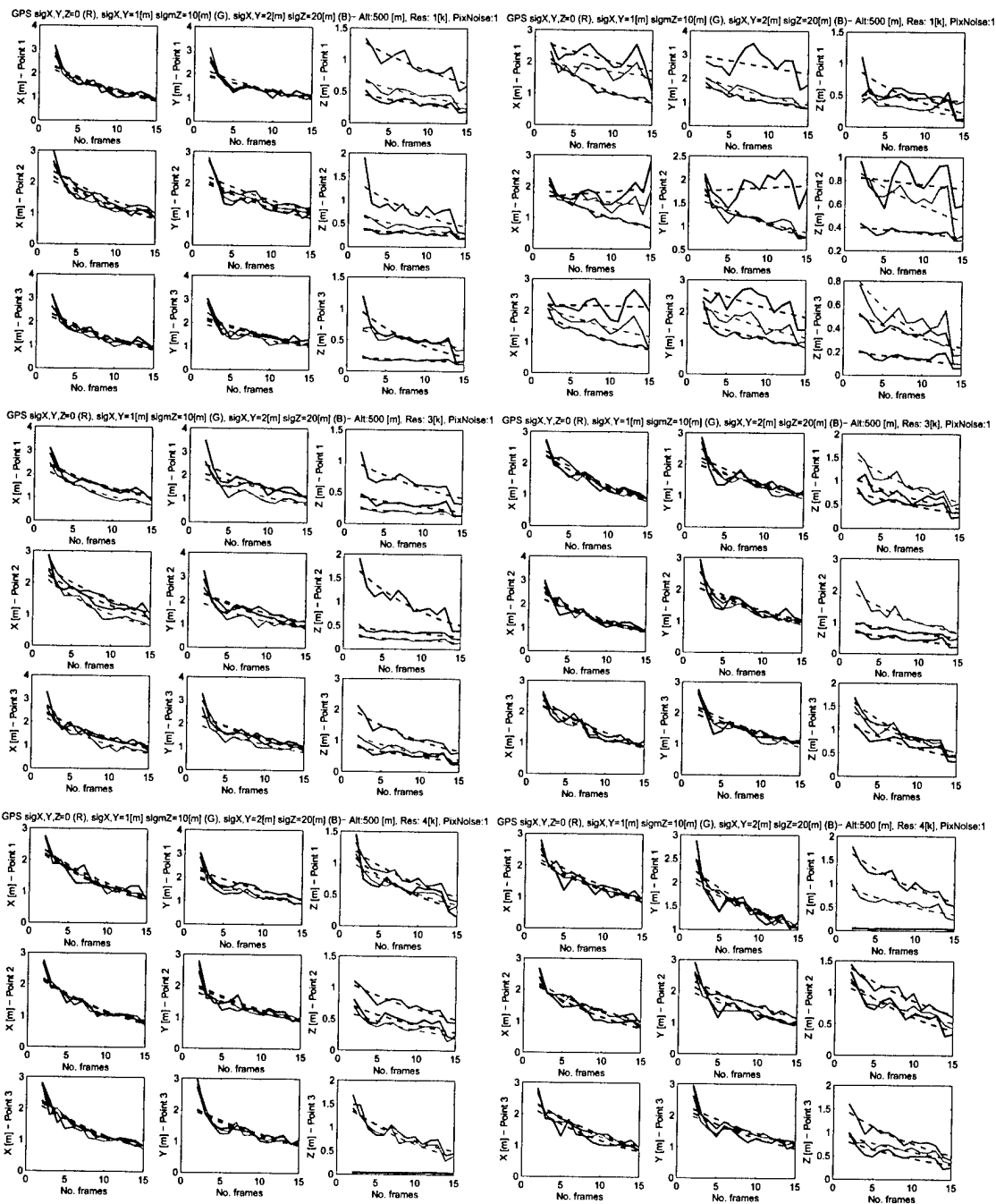


Figure 19: (continued)- Tracking 9 (left) and 12 (right) points.

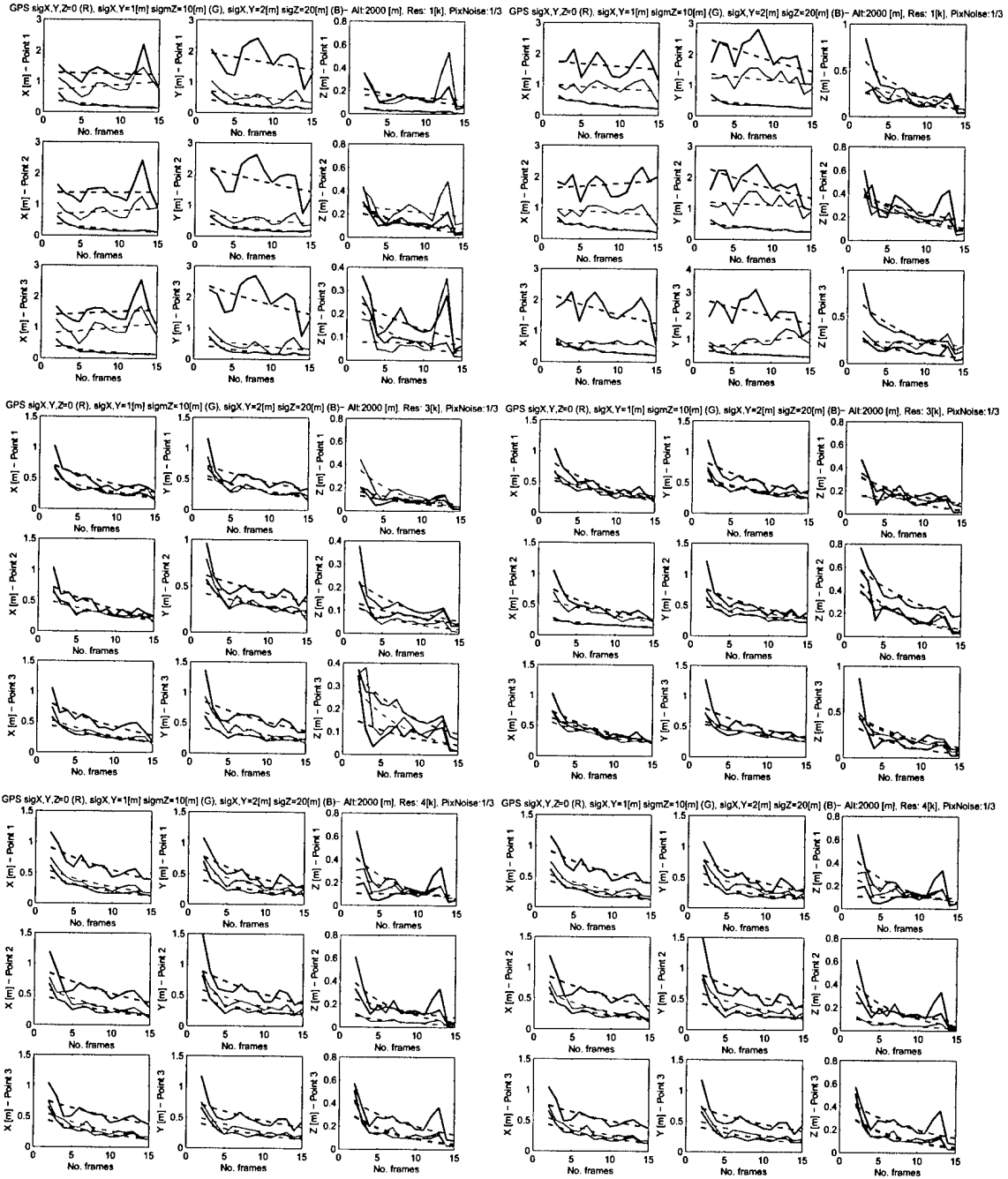


Figure 20: Nonlinear Iterative Solution-- Uncertainty (reconstruction variance [m]) of 3 sample points in terrain feature localization by tracking 3 (left) and 6 (right) points in images with $L \times L$ ($L = \{1, 3, 4\}$) resolutions (Altitude 2000).

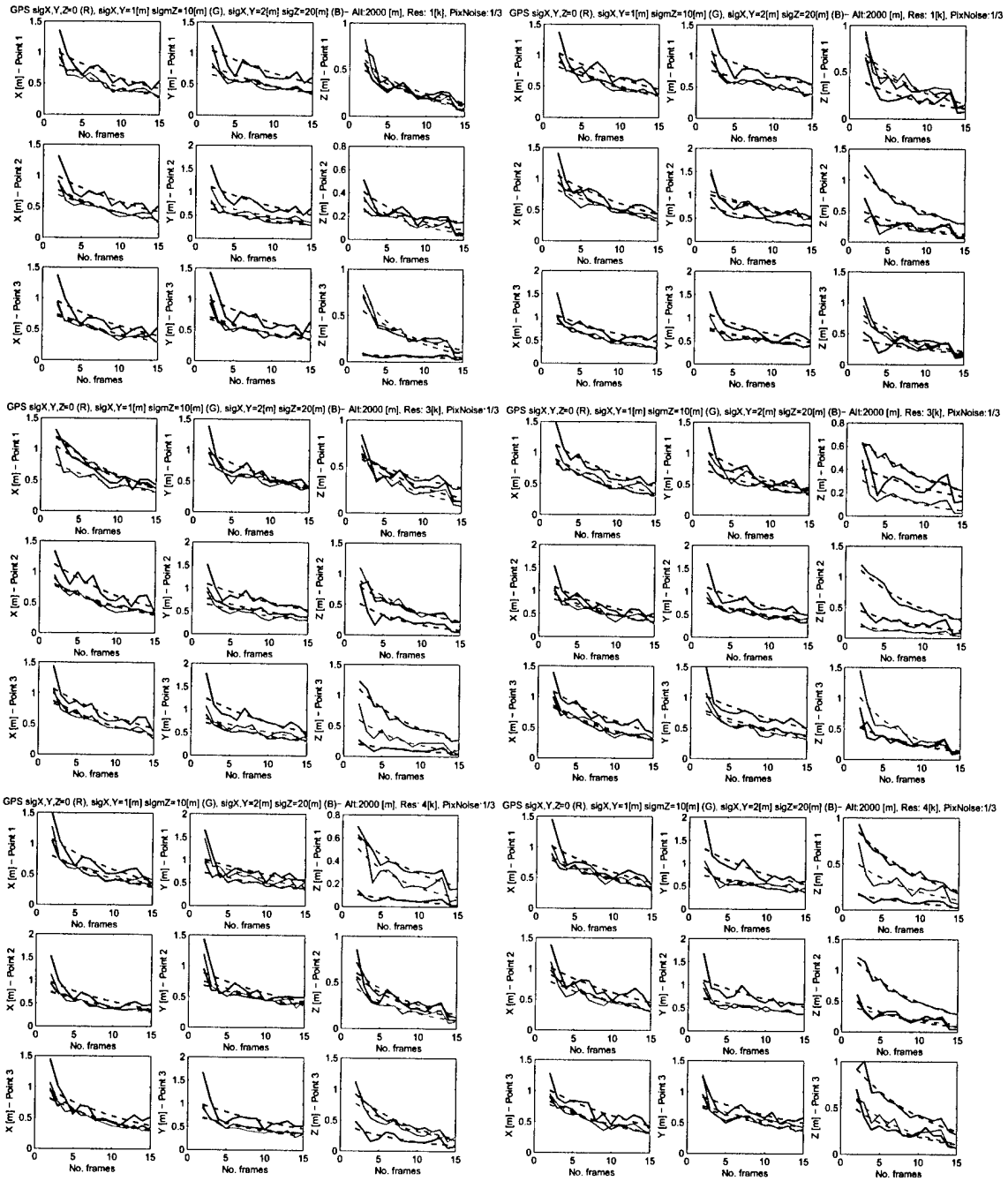


Figure 20: (continued)- Tracking 9 (left) and 12 (right) points.

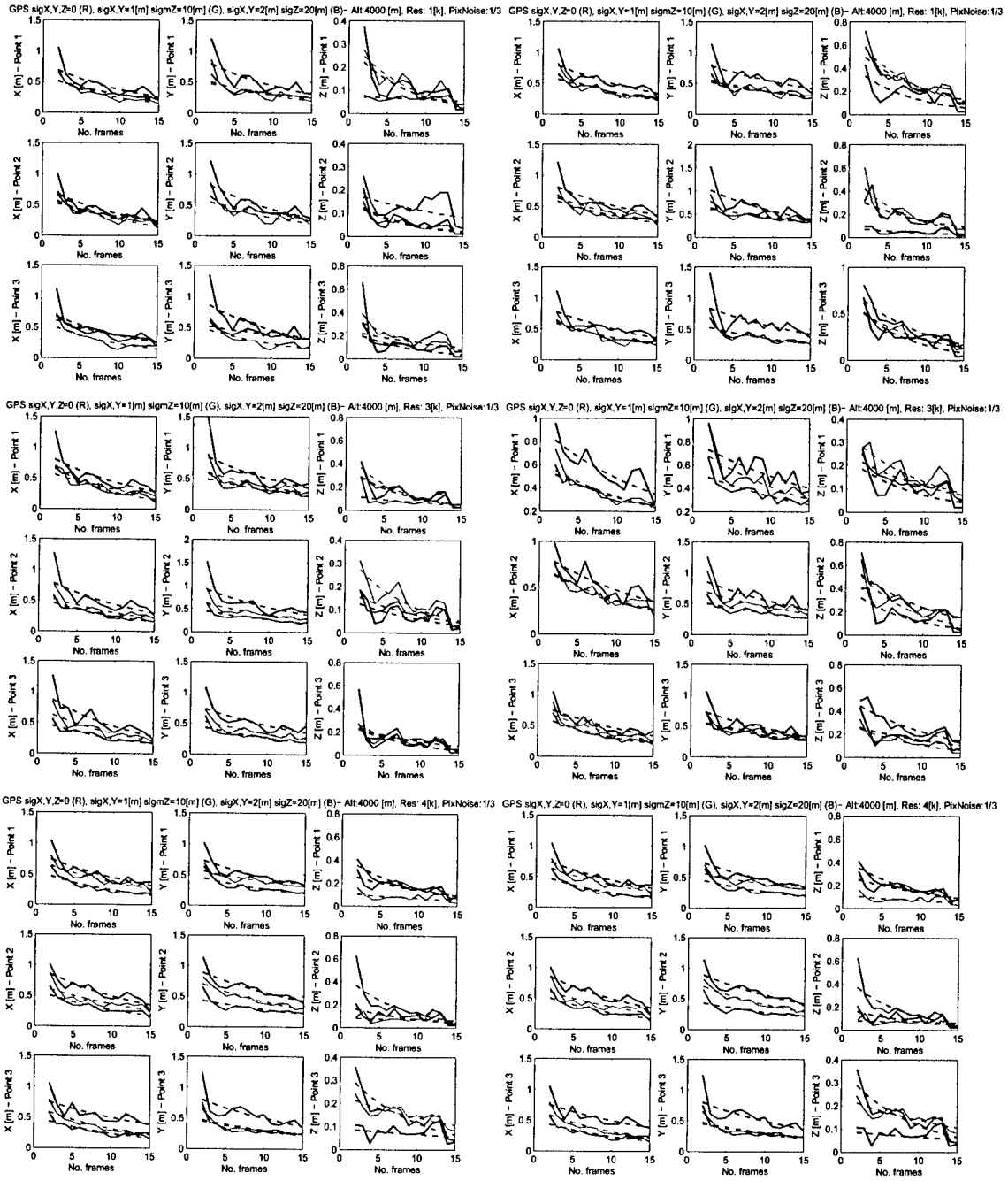


Figure 21: Nonlinear Iterative Solution- Uncertainty (reconstruction variance [m]) of 3 sample points in terrain feature localization by tracking 3 (left) and 6 (right) points in images with $L \times L$ ($L = \{1, 3, 4\}$) resolutions (Altitude 4000).

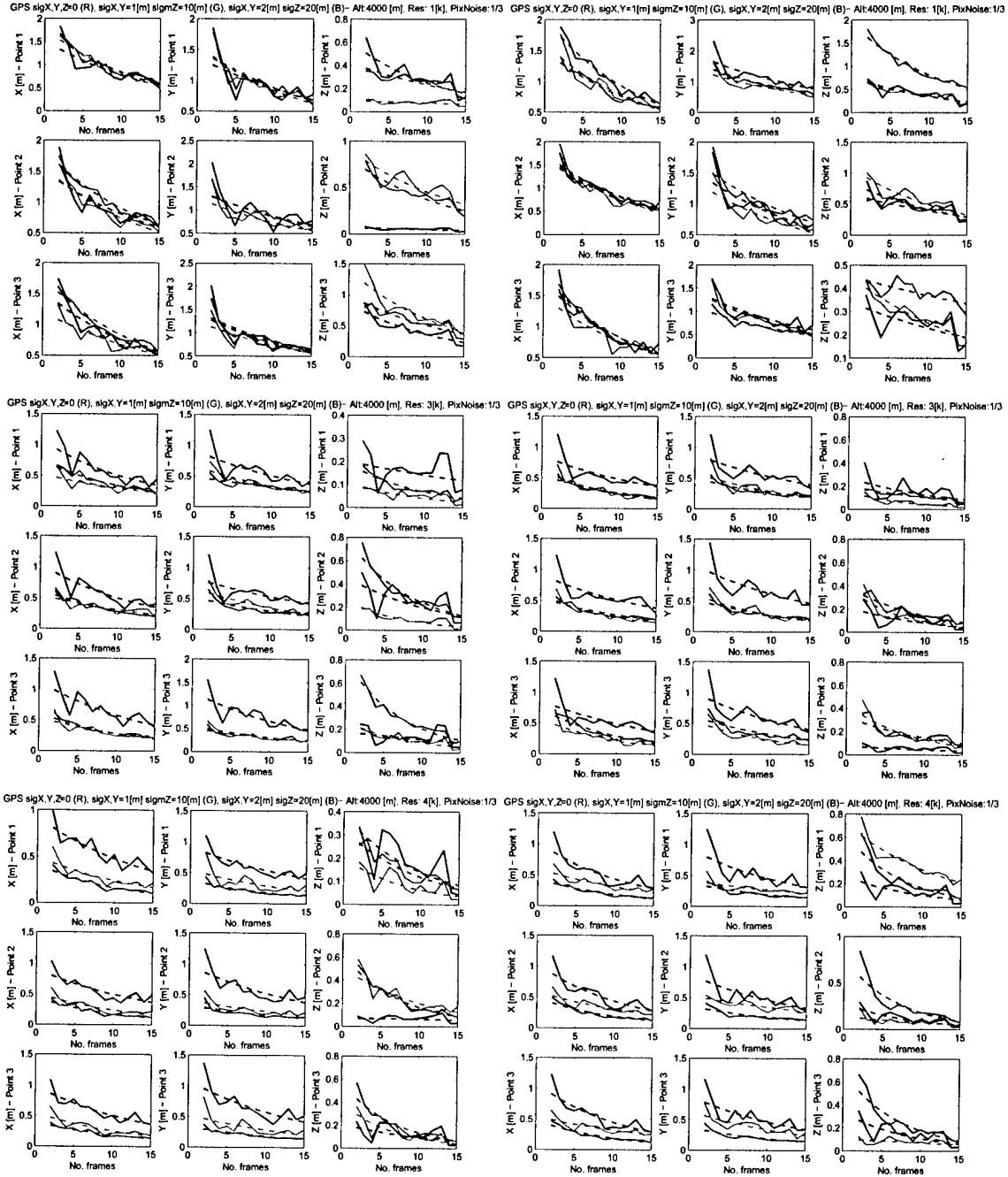


Figure 21: (continued)- Tracking 9 (left) and 12 (right) points.

Appendix 3e: Complete results from nonlinear iterative solution for terrain feature localization

These figures correspond to the same simulations as in the previous appendix, however, they comprise results for all of the terrain feature points.

Each page in this section consists of six— arranged in 3 rows and 2 columns— 3×3 arrays of plots. Each row with 2 such arrays corresponds to one of three $L \times L$ ($L = \{1, 3, 4\}$) image resolutions. Each column of 3 such arrays comprises the results for one case of N ($N = \{3, 6, 9, 12\}$) terrain features being tracked (in $M = \{2 : 15\}$ frames) to determine the 3-D coordinates of landmark terrain features. The complete set for all four values of N are given on two subsequent pages. Each of the 3×3 arrays correspond to error variances of X , Y and Z coordinates of all N terrain features. Each plot has three curves (R,G, and B colors) for GPS variances $\sigma_{GPS} = \{0, 1, 2\}$ [m], respectively. Finally, various pages contain the results for 3 altitudes of 500 [m], 2000 [m], and 4000 [m].

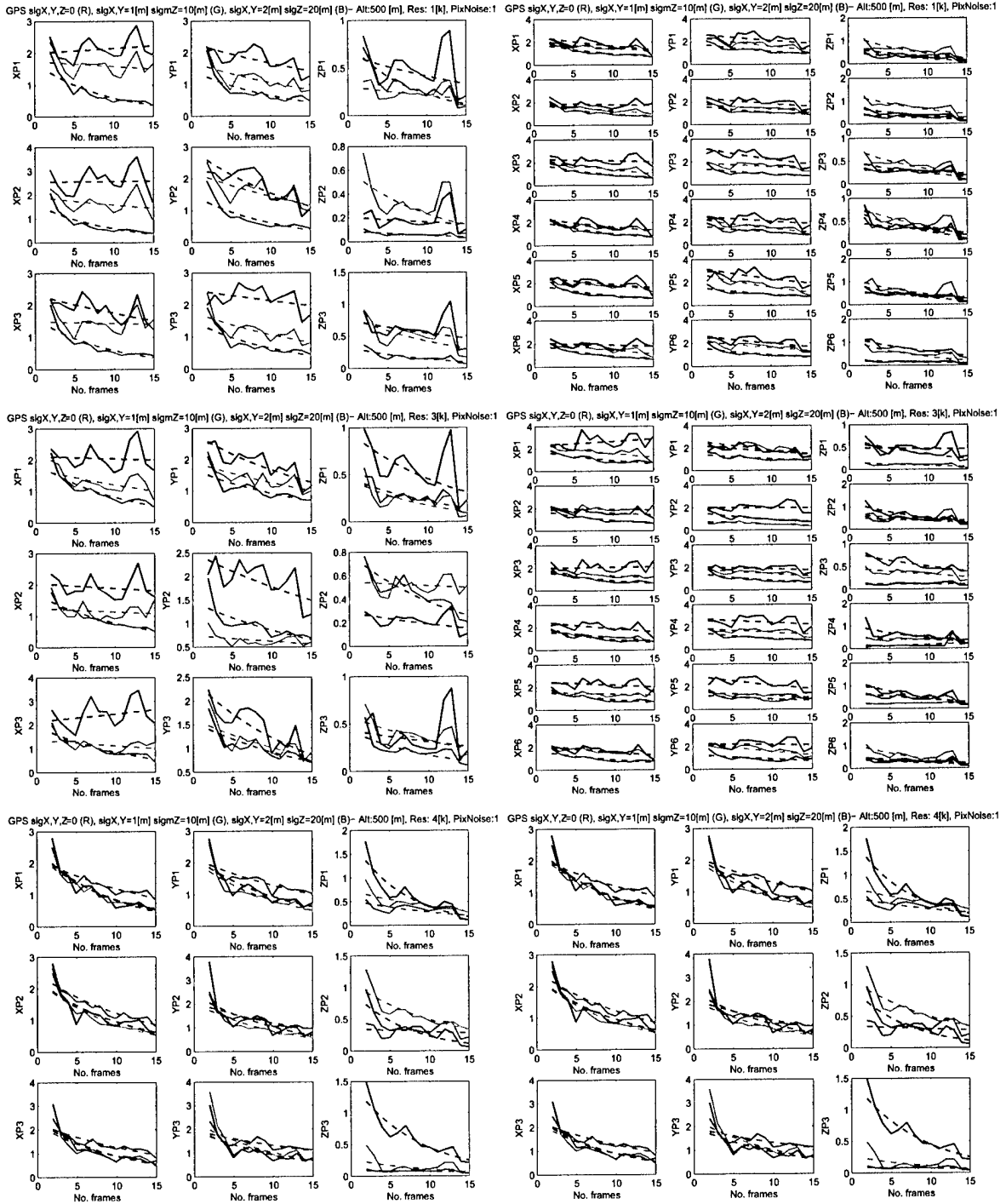


Figure 22: Nonlinear Iterative Solution- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 3 (left) and 6 (right) points in images with $L \times L$ ($L = \{1, 3, 4\}$) resolutions (Altitude 500); See text for details.

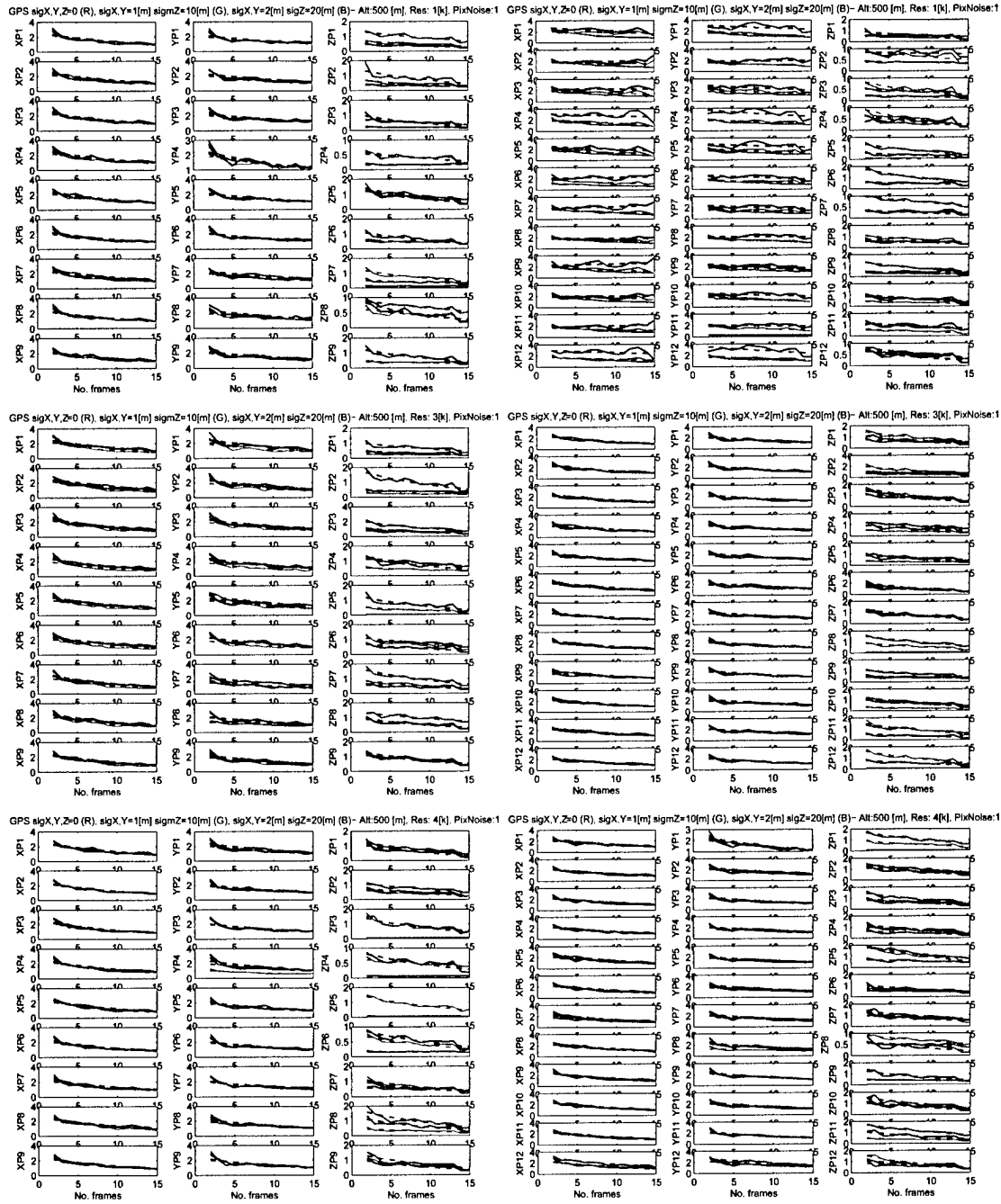


Figure 22: (continued)- Tracking 9 (left) and 12 (right) points.

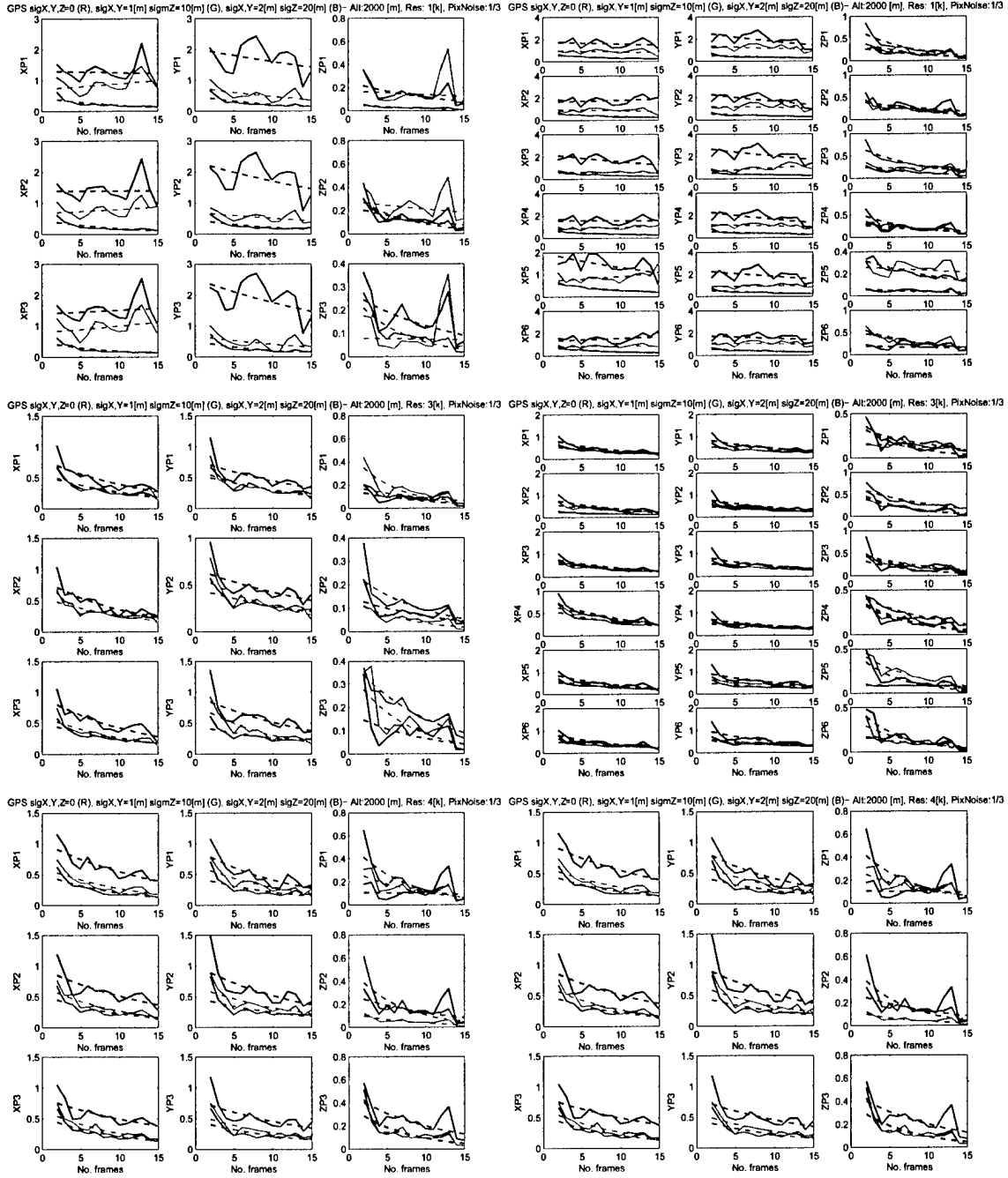


Figure 23: Nonlinear Iterative Solution- Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 3 (left) and 6 (right) points in images with $L \times L$ ($L = \{1, 3, 4\}$) resolutions (Altitude 200); See text for details.

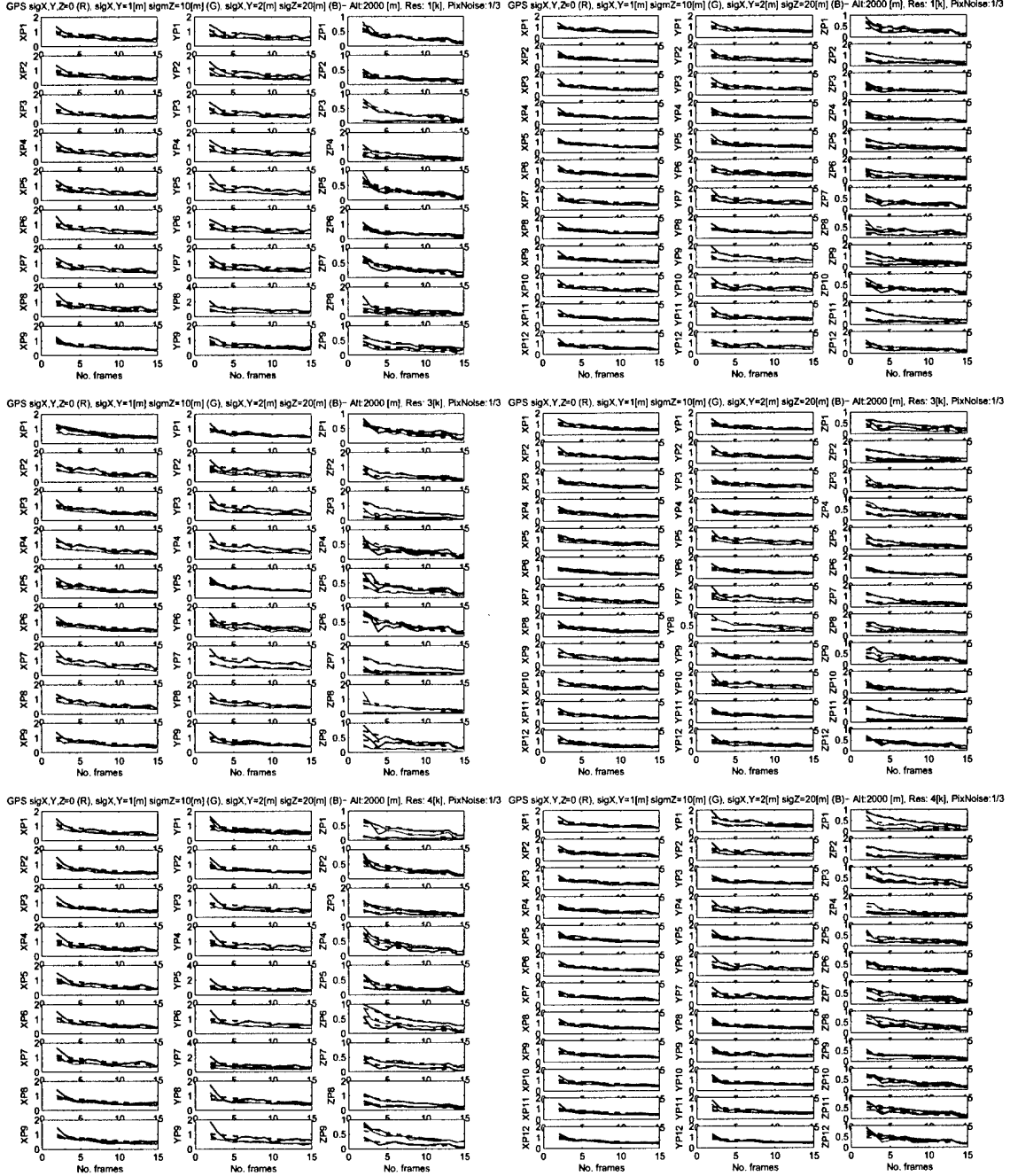


Figure 23: (continued)- Tracking 9 (left) and 12 (right) points.

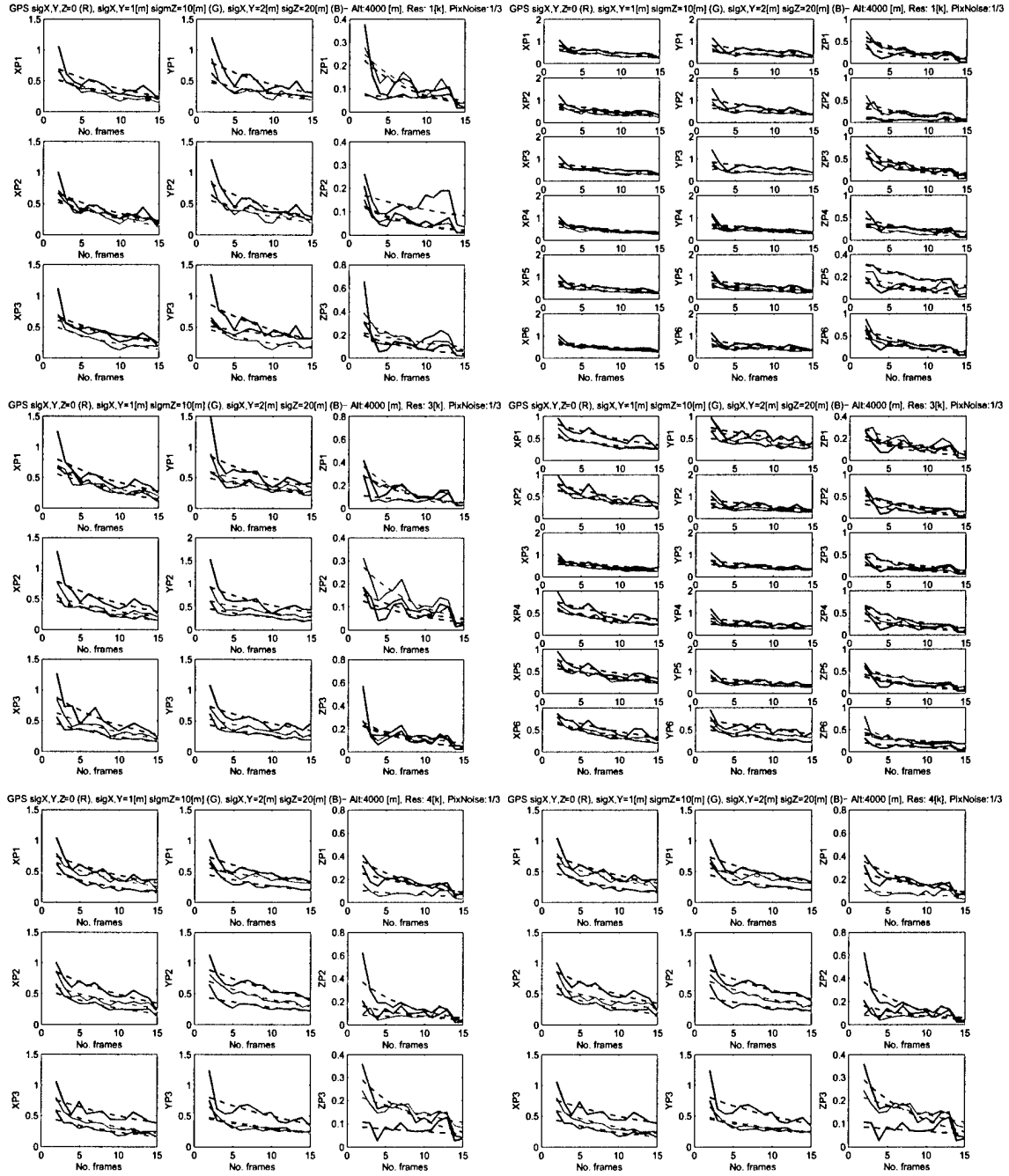


Figure 24: Nonlinear Iterative Solution—Uncertainty (reconstruction variance [m]) in terrain feature localization by tracking 3 (left) and 6 (right) points in images with $L \times L$ ($L = \{1, 3, 4\}$) resolutions (Altitude 500); See text for details.

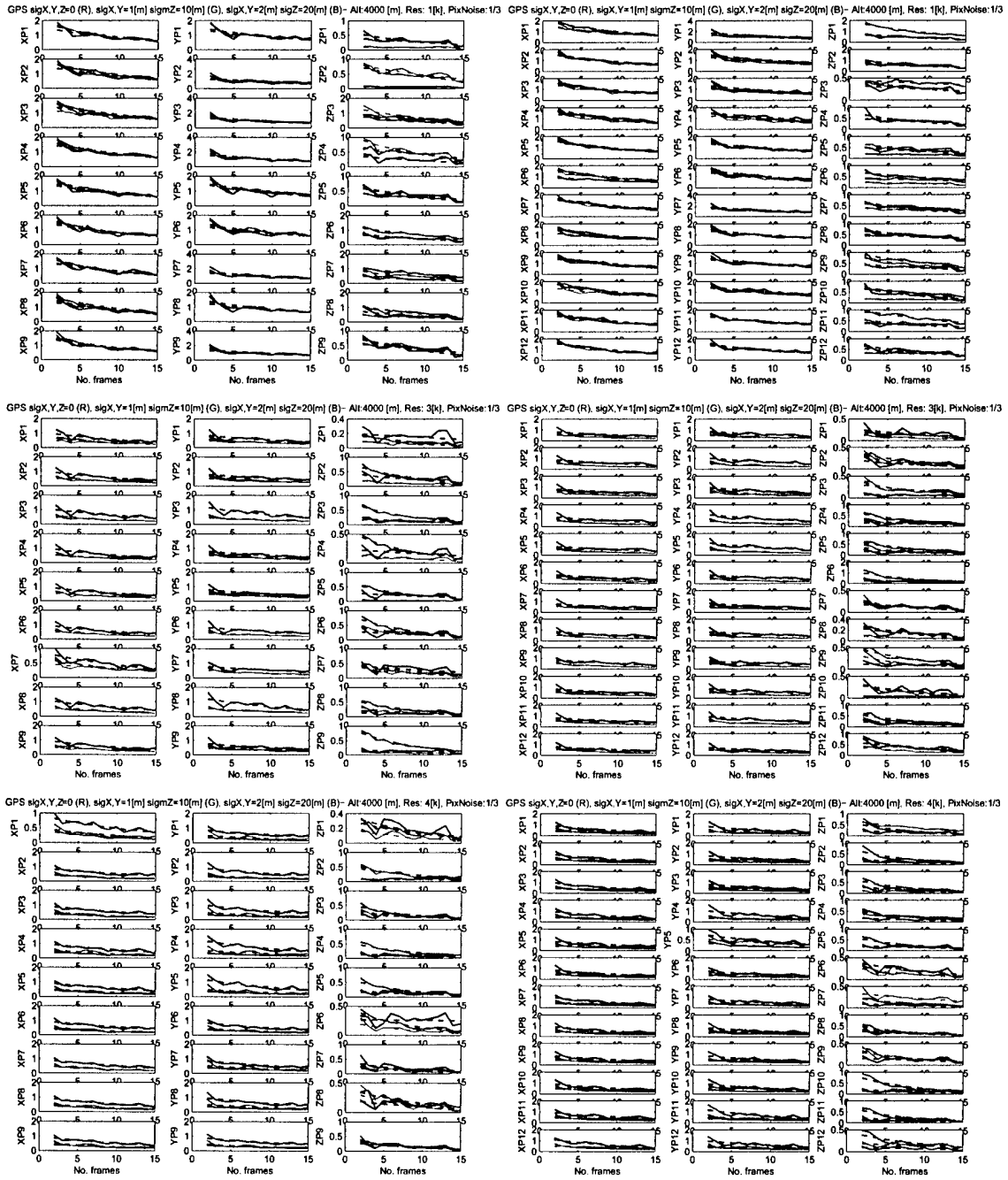


Figure 24: (continued)- Tracking 9 (left) and 12 (right) points.

Appendix 3f: Results of nonlinear iterative solution for UAV pose estimation

This section consists of three 2×3 arrays of plots. Defining the pose as a rotation of the coordinate system with respect to the reference frame, the rows correspond to the variances of the $x - y$ (pitch and roll) and z (heading) components of the rotation vector, as determined from the Rodriguez formula. The three columns are for the 3 camera resolutions. Three curves in each plot are for GPS variances $\sigma_{GPS} = \{0, 1, 2\}$ [m]. Finally, Each array is the result for one of the 3 altitudes– 500 [m], 2000 [m], and 4000 [m].

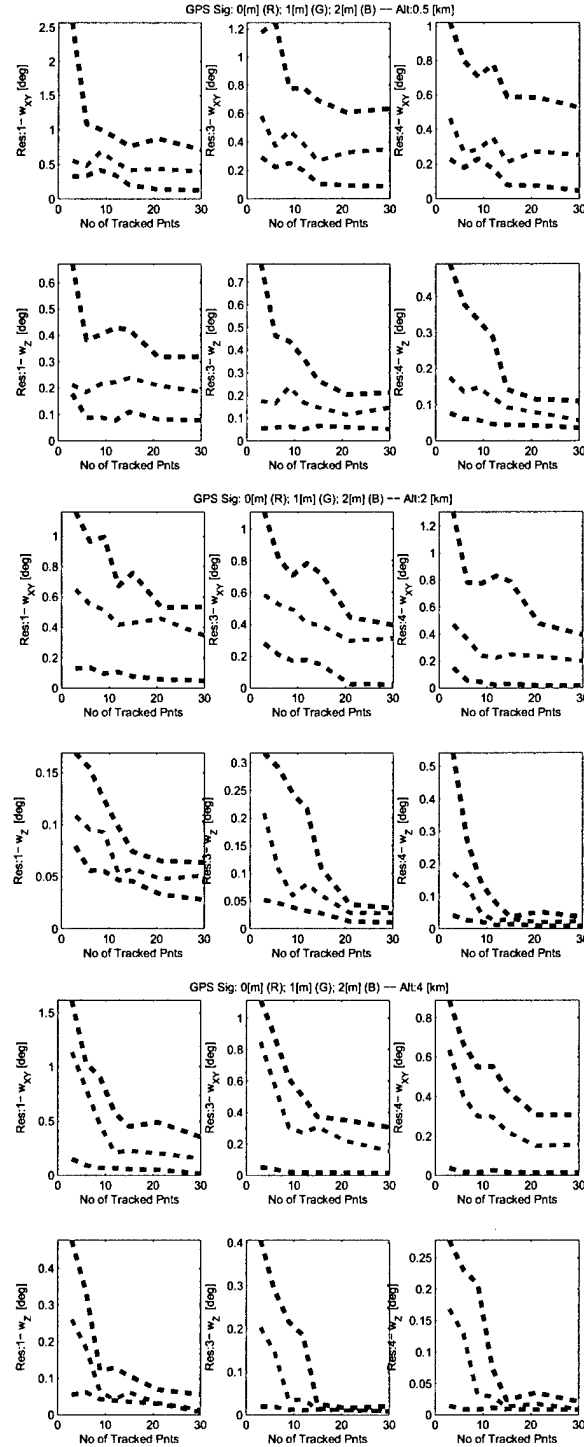


Figure 25: Variances of UAV pose angles— computed from rotation with respect to ref. coordinate system using Rodriguez formula — for various altitudes, GPS measurement uncertainties (R,G,B), and number of terrain feature points tracked in two views. See text for details.